

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
24 October 2002 (24.10.2002)

PCT

(10) International Publication Number  
**WO 02/083736 A2**

- (51) International Patent Classification<sup>7</sup>: **C07K 14/705** (74) Agent: **ZUHN, Donald**; McDonnell Boehnen Hulbert & Berghoff, 300 South Wacker Drive, Chicago, IL 60606 (US).
- (21) International Application Number: **PCT/US02/04397**
- (22) International Filing Date: 14 February 2002 (14.02.2002) (81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data: 60/269,040 14 February 2001 (14.02.2001) US
- (63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application: US 60/269,040 (CON) Filed on 14 February 2001 (14.02.2001)
- (71) Applicant (*for all designated States except US*): **AMGEN, INC.** [US/US]; Amgen, Inc. M/S 27-4-A, One Amgen Center Drive, Thousand Oaks, CA 91320-1799 (US).
- (72) Inventors; and
- (75) Inventors/Applicants (*for US only*): **ELLIOTT, Steven**, G. [US/US]; 1040 Golden Crest Avenue, Newbury Park, CA 91320 (US). **ROGERS, Norma** [US/US]; 4042 Milano Place, Moorpark, CA 93021 (US). **BUSSE, Leigh, Anne** [US/US]; 887 Yearling Court, Camarillo, CA 93010 (US).
- (84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

**Published:**

— *without international search report and to be republished upon receipt of that report*

*For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

(54) Title: **G-PROTEIN COUPLED RECEPTOR MOLECULES AND USES THEREOF**

(57) Abstract: The present invention provides G-Protein Coupled Receptor (GPCR) polypeptides and nucleic acid molecules encoding the same. The invention also provides selective binding agents, vectors, host cells, and methods for producing GPCR polypeptides. The invention further provides pharmaceutical compositions and methods for the diagnosis, treatment, amelioration, and/or prevention of diseases, disorders, and conditions associated with GPCR polypeptides.



**WO 02/083736 A2**

## G-PROTEIN COUPLED RECEPTOR MOLECULES AND USES THEREOF

This application claims the benefit of priority from U.S. Provisional Patent Application No. 60/269,040, filed on February 14, 2001, the disclosure of which is  
5 explicitly incorporated by reference herein.

### Field of the Invention

The present invention relates to G-Protein Coupled Receptor (GPCR) polypeptides and nucleic acid molecules encoding the same. The invention also  
10 relates to selective binding agents, vectors, host cells, and methods for producing GPCR polypeptides. The invention further relates to pharmaceutical compositions and methods for the diagnosis, treatment, amelioration, and/or prevention of diseases, disorders, and conditions associated with GPCR polypeptides.

### Background of the Invention

Technical advances in the identification, cloning, expression, and manipulation of nucleic acid molecules and the deciphering of the human genome have greatly accelerated the discovery of novel therapeutics. Rapid nucleic acid sequencing techniques can now generate sequence information at unprecedented rates  
20 and, coupled with computational analyses, allow the assembly of overlapping sequences into partial and entire genomes and the identification of polypeptide-encoding regions. A comparison of a predicted amino acid sequence against a database compilation of known amino acid sequences allows one to determine the extent of homology to previously identified sequences and/or structural landmarks.  
25 The cloning and expression of a polypeptide-encoding region of a nucleic acid molecule provides a polypeptide product for structural and functional analyses. The manipulation of nucleic acid molecules and encoded polypeptides may confer advantageous properties on a product for use as a therapeutic.

In spite of the significant technical advances in genome research over the past  
30 decade, the potential for the development of novel therapeutics based on the human genome is still largely unrealized. Many genes encoding potentially beneficial polypeptide therapeutics or those encoding polypeptides, which may act as "targets" for therapeutic molecules, have still not been identified. Accordingly, it is an

object of the invention to identify novel polypeptides, and nucleic acid molecules encoding the same, which have diagnostic or therapeutic benefit.

G-protein Coupled Receptors (GPCRs) comprise a large superfamily of integral membrane proteins that are critical constituents of the signal transduction cascade in mammalian cells. The process of signal transduction effected by members of this superfamily involves the binding of an extracellular molecule, such as a hormone or neurotransmitter, to the G-protein coupled receptor, the interaction of the G-protein receptor with a heterotrimeric G-protein, and the subsequent activation of an effector molecule by the G-protein (See Fraser *et al.*, 1994, *Prog. Nucleic Acid Res. Mol. Biol.* 49:113-56). The characteristic features of GPCR polypeptides include seven transmembrane domains of 20-28 hydrophobic residues that span the lipid bilayer in an  $\alpha$ -helical arrangement, six alternating intracellular and extracellular hydrophilic loops connecting the hydrophobic domains, and an extracellular amino-terminal region and intracellular carboxyl-terminal region (See Fraser *et al.*, 1994).

Over the past fifteen years, nearly 350 therapeutic agents targeting G-protein coupled receptors have been successfully introduced onto the market, demonstrating that G-protein coupled receptors have an established, proven history as therapeutic targets for the treatment of human disease. Clearly there is a need in the art for the identification and characterization of other members of this superfamily that may play a role in diagnosing, treating, ameliorating, or preventing diseases or disorders.

#### Summary of the Invention

The present invention relates to novel GPCR nucleic acid molecules and encoded polypeptides.

The invention provides for an isolated nucleic acid molecule comprising:

- (a) the nucleotide sequence as set forth in any of SEQ ID NO: 1, SEQ ID NO: 3, or SEQ ID NO: 5;
  - (b) a nucleotide sequence encoding the polypeptide as set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6;
  - (c) a nucleotide sequence that hybridizes under at least moderately stringent conditions to the complement of the nucleotide sequence of either (a) or (b);
- or

(d) a nucleotide sequence complementary to the nucleotide sequence of either (a) or (b).

The invention also provides for an isolated nucleic acid molecule comprising:

5 (a) a nucleotide sequence encoding a polypeptide that is at least about 70 percent identical to the polypeptide as set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6, wherein the encoded polypeptide has an activity of the polypeptide set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6;

10 (b) a nucleotide sequence encoding an allelic variant or splice variant of the nucleotide sequence as set forth in any of SEQ ID NO: 1, SEQ ID NO: 3, or SEQ ID NO: 5 or the nucleotide sequence of (a);

(c) a region of the nucleotide sequence of any of SEQ ID NO: 1, SEQ ID NO: 3, or SEQ ID NO: 5 or the nucleotide sequence of (a) or (b) encoding a polypeptide fragment of at least about 25 amino acid residues, wherein the  
15 polypeptide fragment has an activity of the polypeptide set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6, or is antigenic;

(d) a region of the nucleotide sequence of any of SEQ ID NO: 1, SEQ ID NO: 3, or SEQ ID NO: 5 or the nucleotide sequence of any of (a) - (c) comprising a fragment of at least about 16 nucleotides;

20 (e) a nucleotide sequence that hybridizes under at least moderately stringent conditions to the complement of the nucleotide sequence of any of (a) - (d);  
or

(f) a nucleotide sequence complementary to the nucleotide sequence of any of (a) - (d).

25

The invention further provides for an isolated nucleic acid molecule comprising:

(a) a nucleotide sequence encoding a polypeptide as set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6 with at least one conservative  
30 amino acid substitution, wherein the encoded polypeptide has an activity of the polypeptide set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6;

(b) a nucleotide sequence encoding a polypeptide as set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6 with at least one amino acid



insertion, wherein the encoded polypeptide has an activity of the polypeptide set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6;

(c) a nucleotide sequence encoding a polypeptide as set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6 with at least one amino acid deletion, wherein the encoded polypeptide has an activity of the polypeptide set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6;

(d) a nucleotide sequence encoding a polypeptide as set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6 that has a C- and/or N- terminal truncation, wherein the encoded polypeptide has an activity of the polypeptide set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6;

(e) a nucleotide sequence encoding a polypeptide as set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6 with at least one modification that is an amino acid substitution, an amino acid insertion, an amino acid deletion, C-terminal truncation, or N-terminal truncation, wherein the encoded polypeptide has an activity of the polypeptide set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6;

(f) a nucleotide sequence of any of (a) - (e) comprising a fragment of at least about 16 nucleotides;

(g) a nucleotide sequence that hybridizes under at least moderately stringent conditions to the complement of the nucleotide sequence of any of (a) - (f);  
or

(h) a nucleotide sequence complementary to the nucleotide sequence of any of (a) - (e).

The present invention provides for an isolated polypeptide comprising the amino acid as set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6.

The invention also provides for an isolated polypeptide comprising:

(a) an amino acid sequence for an ortholog of any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6;

(b) an amino acid sequence that is at least about 70 percent identical to the amino acid sequence of any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6, wherein the polypeptide has an activity of the polypeptide set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6;

(c) a fragment of the amino acid sequence set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6 comprising at least about 25 amino acid residues, wherein the fragment has an activity of the polypeptide set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6, or is antigenic; or

5 (d) an amino acid sequence for an allelic variant or splice variant of the amino acid sequence as set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6 or the amino acid sequence of either (a) or (b).

The invention further provides for an isolated polypeptide comprising:

10 (a) the amino acid sequence as set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6 with at least one conservative amino acid substitution, wherein the polypeptide has an activity of the polypeptide set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6;

(b) the amino acid sequence as set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6 with at least one amino acid insertion, wherein the polypeptide has an activity of the polypeptide set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6;

(c) the amino acid sequence as set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6 with at least one amino acid deletion, wherein the polypeptide has an activity of the polypeptide set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6;

(d) the amino acid sequence as set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6 that has a C- and/or N- terminal truncation, wherein the polypeptide has an activity of the polypeptide set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6; or

(e) the amino acid sequence as set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6 with at least one modification that is an amino acid substitution, an amino acid insertion, an amino acid deletion, C-terminal truncation, or N-terminal truncation, wherein the polypeptide has an activity of the polypeptide set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6.

The invention still further provides for an isolated polypeptide comprising the amino acid sequence as set forth in SEQ ID NO: 2 with at least one conservative amino acid substitution that is a aspartic acid at position 2; leucine at position 8;

glutamic acid at position 12; proline at position 13; threonine at position 15; leucine at position 24; valine at position 25; leucine at position 27; isoleucine or leucine at position 35; serine at position 49; isoleucine at position 52; isoleucine at position 82; leucine at position 83; alanine at position 87; leucine at position 90; valine at position 91; lysine or methionine at position 94; valine at position 111; methionine at position 123; alanine at position 126; N at position 129; threonine at position 131; alanine at position 134; threonine at position 135; alanine at position 136; valine at position 138; threonine at position 141; methionine at position 152; serine at position 154; arginine at position 159; glycine at position 160; methionine at position 161; leucine or valine at position 162; serine at position 163; valine at position 179; leucine at position 187; threonine at position 190; valine at position 197; asparagine at position 198; valine at position 199; glutamine at position 205; threonine at position 210; arginine at position 216; arginine at position 217; serine at position 227; leucine at position 238; threonine at position 249; threonine or isoleucine at position 258; valine at position 262; leucine at position 266; leucine at position 269; leucine at position 285; alanine at position 290; threonine at position 293; arginine at position 295; arginine at position 300; arginine at position 301; arginine at position 304; threonine at position 305; glutamine at position 306; alanine at position 307; arginine at position 308; serine at position 310; glycine at position 319; serine at position 320; lysine at position 321; serine at position 322; threonine at position 324; aspartic acid at position 325; glycine at position 326; valine at position 327; arginine at position 329; serine at position 330; arginine at position 332; proline at position 334; glycine at position 339; leucine at position 340; glutamine at position 341; and valine at position 342; wherein the polypeptide has an activity of the polypeptide set forth in SEQ ID NO: 2.

25

Also provided are fusion polypeptides comprising GPCR amino acid sequences.

The present invention also provides for an expression vector comprising the isolated nucleic acid molecules as set forth herein, recombinant host cells comprising the recombinant nucleic acid molecules as set forth herein, and a method of producing a GPCR polypeptide comprising culturing the host cells and optionally isolating the polypeptide so produced. Isolation of the expressed polypeptide is described as optional because there may be instances where it is desired to express the polypeptide

on the cell surface or on a cell membrane for use in screening methods for the identification of agonists or antagonists of GPCR activity.

A transgenic non-human animal comprising a nucleic acid molecule encoding a GPCR polypeptide is also encompassed by the invention. The GPCR nucleic acid molecules are introduced into the animal in a manner that allows expression and increased levels of a GPCR polypeptide, which may include increased circulating levels. Alternatively, the GPCR nucleic acid molecules are introduced into the animal in a manner that prevents expression of endogenous GPCR polypeptide (*i.e.*, generates a transgenic animal possessing a GPCR polypeptide gene knockout). The transgenic non-human animal is preferably a mammal, and more preferably a rodent, such as a rat or a mouse.

Also provided are derivatives of the GPCR polypeptides of the present invention.

Additionally provided are selective binding agents such as antibodies and peptides capable of specifically binding the GPCR polypeptides of the invention. Such antibodies and peptides may be agonistic or antagonistic.

Pharmaceutical compositions comprising the nucleotides, polypeptides, or selective binding agents of the invention and one or more pharmaceutically acceptable formulation agents are also encompassed by the invention. The pharmaceutical compositions are used to provide therapeutically effective amounts of the nucleotides or polypeptides of the present invention. The invention is also directed to methods of using the polypeptides, nucleic acid molecules, and selective binding agents.

The GPCR polypeptides and nucleic acid molecules of the present invention may be used to treat, prevent, ameliorate, and/or detect diseases and disorders, including those recited herein.

The present invention also provides a method of assaying test molecules to identify a test molecule that binds to a GPCR polypeptide. The method comprises contacting a GPCR polypeptide with a test molecule to determine the extent of binding of the test molecule to the polypeptide. The method further comprises determining whether such test molecules are agonists or antagonists of a GPCR polypeptide. The present invention further provides a method of testing the impact of molecules on the expression of GPCR polypeptide or on the activity of GPCR polypeptide.

Methods of regulating expression and modulating (*i.e.*, increasing or decreasing) levels of a GPCR polypeptide are also encompassed by the invention. One method comprises administering to an animal a nucleic acid molecule encoding a GPCR polypeptide. In another method, a nucleic acid molecule comprising elements  
5 that regulate or modulate the expression of a GPCR polypeptide may be administered. Examples of these methods include gene therapy, cell therapy, and anti-sense therapy as further described herein.

GPCR polypeptides can be used for identifying ligands thereof. Various forms of "expression cloning" have been used for cloning ligands for receptors (*See*,  
10 *e.g.*, Davis *et al.*, 1996, *Cell*, 87:1161-69). These and other GPCR ligand cloning experiments are described in greater detail herein. Isolation of the GPCR ligand(s) allows for the identification or development of novel agonists and/or antagonists of the GPCR signaling pathway. Such agonists and antagonists include GPCR ligand(s), anti-GPCR ligand antibodies and derivatives thereof, small molecules, or antisense  
15 oligonucleotides, any of which can be used for potentially treating one or more diseases or disorders, including those recited herein.

#### Brief Description of the Figures

Figures 1A-1B illustrate the nucleotide sequence of the human GPCR gene (SEQ ID  
20 NO: 1) and the deduced amino acid sequence of human GPCR polypeptide (SEQ ID NO: 2);

Figures 2A-2D illustrate the nucleotide sequence of the murine GPCR gene (SEQ ID  
NO: 3) and the deduced amino acid sequence of murine GPCR polypeptide (SEQ ID  
25 NO: 4);

Figures 3A-3C illustrate the nucleotide sequence of the rat GPCR gene (SEQ ID NO:  
5) and the deduced amino acid sequence of rat GPCR polypeptide (SEQ ID NO: 6);

30 Figure 4 illustrates the amino acid sequence alignment of human GPCR polypeptide (hu\_GPCR; SEQ ID NO: 2), murine GPCR polypeptide (mu\_GPCR; SEQ ID NO: 4) and rat GPCR polypeptide (ra\_GPCR; SEQ ID NO: 6). Putative transmembrane domains are indicated (underline);

Figure 5 illustrates the level of GPCR mRNA expression as determined in several human tissues by quantitative PCR.

#### Detailed Description of the Invention

5           The section headings used herein are for organizational purposes only and are not to be construed as limiting the subject matter described. All references cited in this application are expressly incorporated by reference herein.

#### Definitions

10           The terms "GPCR gene" or "GPCR nucleic acid molecule" or "GPCR polynucleotide" refer to a nucleic acid molecule comprising or consisting of a nucleotide sequence as set forth in any of SEQ ID NO: 1, SEQ ID NO: 3, or SEQ ID NO: 5, a nucleotide sequence encoding the polypeptide as set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6, and nucleic acid molecules as defined  
15   herein.

          The term "GPCR polypeptide allelic variant" refers to one of several possible naturally occurring alternate forms of a gene occupying a given locus on a chromosome of an organism or a population of organisms.

          The term "GPCR polypeptide splice variant" refers to a nucleic acid molecule,  
20   usually RNA, which is generated by alternative processing of intron sequences in an RNA transcript of GPCR polypeptide amino acid sequence as set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6.

          The term "isolated nucleic acid molecule" refers to a nucleic acid molecule of the invention that (1) has been separated from at least about 50 percent of proteins,  
25   lipids, carbohydrates, or other materials with which it is naturally found when total nucleic acid is isolated from the source cells, (2) is not linked to all or a portion of a polynucleotide to which the "isolated nucleic acid molecule" is linked in nature, (3) is operably linked to a polynucleotide which it is not linked to in nature, or (4) does not occur in nature as part of a larger polynucleotide sequence. Preferably, the isolated  
30   nucleic acid molecule of the present invention is substantially free from any other contaminating nucleic acid molecule(s) or other contaminants that are found in its natural environment that would interfere with its use in polypeptide production or its therapeutic, diagnostic, prophylactic or research use.

The term "nucleic acid sequence" or "nucleic acid molecule" refers to a DNA or RNA sequence. The term encompasses molecules formed from any of the known base analogs of DNA and RNA such as, but not limited to 4-acetylcytosine, 8-hydroxy-N6-methyladenosine, aziridinyl-cytosine, pseudoisocytosine, 5-  
5 (carboxyhydroxymethyl) uracil, 5-fluorouracil, 5-bromouracil, 5-carboxymethylaminomethyl-2-thiouracil, 5-carboxy-methylaminomethyluracil, dihydrouracil, inosine, N6-iso-pentenyladenine, 1-methyladenine, 1-methylpseudouracil, 1-methylguanine, 1-methylinosine, 2,2-dimethyl-guanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-  
10 methyladenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyamino-methyl-2-thiouracil, beta-D-mannosylqueosine, 5' -methoxycarbonyl-methyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid, oxybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, N-  
15 uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid, pseudouracil, queosine, 2-thiocytosine, and 2,6-diaminopurine.

The term "vector" is used to refer to any molecule (*e.g.*, nucleic acid, plasmid, or virus) used to transfer coding information to a host cell.

The term "expression vector" refers to a vector that is suitable for  
20 transformation of a host cell and contains nucleic acid sequences that direct and/or control the expression of inserted heterologous nucleic acid sequences. Expression includes, but is not limited to, processes such as transcription, translation, and RNA splicing, if introns are present.

The term "operably linked" is used herein to refer to an arrangement of  
25 flanking sequences wherein the flanking sequences so described are configured or assembled so as to perform their usual function. Thus, a flanking sequence operably linked to a coding sequence may be capable of effecting the replication, transcription and/or translation of the coding sequence. For example, a coding sequence is operably linked to a promoter when the promoter is capable of directing transcription  
30 of that coding sequence. A flanking sequence need not be contiguous with the coding sequence, so long as it functions correctly. Thus, for example, intervening untranslated yet transcribed sequences can be present between a promoter sequence and the coding sequence and the promoter sequence can still be considered "operably linked" to the coding sequence.

The term "host cell" is used to refer to a cell which has been transformed, or is capable of being transformed with a nucleic acid sequence and then of expressing a selected gene of interest. The term includes the progeny of the parent cell, whether or not the progeny is identical in morphology or in genetic make-up to the original parent, so long as the selected gene is present.

The term "GPCR polypeptide" refers to a polypeptide comprising the amino acid sequence of any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6 and related polypeptides. Related polypeptides include GPCR polypeptide fragments, GPCR polypeptide orthologs, GPCR polypeptide variants, and GPCR polypeptide derivatives, which possess at least one activity of the polypeptide as set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6. GPCR polypeptides may be mature polypeptides, as defined herein, and may or may not have an amino-terminal methionine residue, depending on the method by which they are prepared.

The term "GPCR polypeptide fragment" refers to a polypeptide that comprises a truncation at the amino-terminus (with or without a leader sequence) and/or a truncation at the carboxyl-terminus of the polypeptide as set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6. The term "GPCR polypeptide fragment" also refers to amino-terminal and/or carboxyl-terminal truncations of GPCR polypeptide orthologs, GPCR polypeptide derivatives, or GPCR polypeptide variants, or to amino-terminal and/or carboxyl-terminal truncations of the polypeptides encoded by GPCR polypeptide allelic variants or GPCR polypeptide splice variants. GPCR polypeptide fragments may result from alternative RNA splicing or from *in vivo* protease activity. Membrane-bound forms of a GPCR polypeptide are also contemplated by the present invention. In preferred embodiments, truncations and/or deletions comprise about 10 amino acids, or about 20 amino acids, or about 50 amino acids, or about 75 amino acids, or about 100 amino acids, or more than about 100 amino acids. The polypeptide fragments so produced will comprise about 25 contiguous amino acids, or about 50 amino acids, or about 75 amino acids, or about 100 amino acids, or about 150 amino acids, or about 200 amino acids, or more than about 200 amino acids. Such GPCR polypeptide fragments may optionally comprise an amino-terminal methionine residue. It will be appreciated that such fragments can be used, for example, to generate antibodies to GPCR polypeptides.

The term "GPCR polypeptide ortholog" refers to a polypeptide from another species that corresponds to GPCR polypeptide amino acid sequence as set forth in any



of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6. For example, mouse and human GPCR polypeptides are considered orthologs of each other.

The term "GPCR polypeptide variants" refers to GPCR polypeptides comprising amino acid sequences having one or more amino acid sequence  
5 substitutions, deletions (such as internal deletions and/or GPCR polypeptide fragments), and/or additions (such as internal additions and/or GPCR fusion polypeptides) as compared to the GPCR polypeptide amino acid sequence set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6 (with or without a leader sequence). Variants may be naturally occurring (*e.g.*, GPCR polypeptide allelic  
10 variants, GPCR polypeptide orthologs, and GPCR polypeptide splice variants) or artificially constructed. Such GPCR polypeptide variants may be prepared from the corresponding nucleic acid molecules having a DNA sequence that varies accordingly from the DNA sequence as set forth in any of SEQ ID NO: 1, SEQ ID NO: 3, or SEQ ID NO: 5. In preferred embodiments, the variants have from 1 to 3, or from 1 to 5, or  
15 from 1 to 10, or from 1 to 15, or from 1 to 20, or from 1 to 25, or from 1 to 50, or from 1 to 75, or from 1 to 100, or more than 100 amino acid substitutions, insertions, additions and/or deletions, wherein the substitutions may be conservative, or non-conservative, or any combination thereof.

The term "GPCR polypeptide derivatives" refers to the polypeptide as set  
20 forth in any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6, GPCR polypeptide fragments, GPCR polypeptide orthologs, or GPCR polypeptide variants, as defined herein, that have been chemically modified. The term "GPCR polypeptide derivatives" also refers to the polypeptides encoded by GPCR polypeptide allelic variants or GPCR polypeptide splice variants, as defined herein, that have been  
25 chemically modified.

The term "mature GPCR polypeptide" refers to a GPCR polypeptide lacking a leader sequence. A mature GPCR polypeptide may also include other modifications such as proteolytic processing of the amino-terminus (with or without a leader sequence) and/or the carboxyl-terminus, cleavage of a smaller polypeptide from a  
30 larger precursor, N-linked and/or O-linked glycosylation, and the like.

The term "GPCR fusion polypeptide" refers to a fusion of one or more amino acids (such as a heterologous protein or peptide) at the amino- or carboxyl-terminus of the polypeptide as set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6, GPCR polypeptide fragments, GPCR polypeptide orthologs, GPCR

polypeptide variants, or GPCR derivatives, as defined herein. The term "GPCR fusion polypeptide" also refers to a fusion of one or more amino acids at the amino- or carboxyl-terminus of the polypeptide encoded by GPCR polypeptide allelic variants or GPCR polypeptide splice variants, as defined herein.

5           The term "biologically active GPCR polypeptides" refers to GPCR polypeptides having at least one activity characteristic of the polypeptide comprising the amino acid sequence of any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6. In addition, a GPCR polypeptide may be active as an immunogen; that is, the GPCR polypeptide contains at least one epitope to which antibodies may be raised.

10           The term "isolated polypeptide" refers to a polypeptide of the present invention that (1) has been separated from at least about 50 percent of polynucleotides, lipids, carbohydrates, or other materials with which it is naturally found when isolated from the source cell, (2) is not linked (by covalent or noncovalent interaction) to all or a portion of a polypeptide to which the "isolated  
15   polypeptide" is linked in nature, (3) is operably linked (by covalent or noncovalent interaction) to a polypeptide with which it is not linked in nature, or (4) does not occur in nature. Preferably, the isolated polypeptide is substantially free from any other contaminating polypeptides or other contaminants that are found in its natural environment that would interfere with its therapeutic, diagnostic, prophylactic or  
20   research use.

          The term "identity," as known in the art, refers to a relationship between the sequences of two or more polypeptide molecules or two or more nucleic acid molecules, as determined by comparing the sequences. In the art, "identity" also means the degree of sequence relatedness between nucleic acid molecules or  
25   polypeptides, as the case may be, as determined by the match between strings of two or more nucleotide or two or more amino acid sequences. "Identity" measures the percent of identical matches between the smaller of two or more sequences with gap alignments (if any) addressed by a particular mathematical model or computer program (*i.e.*, "algorithms").

30           The term "similarity" is a related concept, but in contrast to "identity," "similarity" refers to a measure of relatedness that includes both identical matches and conservative substitution matches. If two polypeptide sequences have, for example, 10/20 identical amino acids, and the remainder are all non-conservative substitutions, then the percent identity and similarity would both be 50%. If in the same example,

there are five more positions where there are conservative substitutions, then the percent identity remains 50%, but the percent similarity would be 75% (15/20). Therefore, in cases where there are conservative substitutions, the percent similarity between two polypeptides will be higher than the percent identity between those two polypeptides.

The term "naturally occurring" or "native" when used in connection with biological materials such as nucleic acid molecules, polypeptides, host cells, and the like, refers to materials which are found in nature and are not manipulated by man. Similarly, "non-naturally occurring" or "non-native" as used herein refers to a material that is not found in nature or that has been structurally modified or synthesized by man.

The terms "effective amount" and "therapeutically effective amount" each refer to the amount of a GPCR polypeptide or GPCR nucleic acid molecule used to support an observable level of one or more biological activities of the GPCR polypeptides as set forth herein.

The term "pharmaceutically acceptable carrier" or "physiologically acceptable carrier" as used herein refers to one or more formulation materials suitable for accomplishing or enhancing the delivery of the GPCR polypeptide, GPCR nucleic acid molecule, or GPCR selective binding agent as a pharmaceutical composition.

The term "antigen" refers to a molecule or a portion of a molecule capable of being bound by a selective binding agent, such as an antibody, and additionally capable of being used in an animal to produce antibodies capable of binding to an epitope of that antigen. An antigen may have one or more epitopes.

The term "selective binding agent" refers to a molecule or molecules having specificity for a GPCR polypeptide. As used herein, the terms, "specific" and "specificity" refer to the ability of the selective binding agents to bind to human GPCR polypeptides and not to bind to human non-GPCR polypeptides. It will be appreciated, however, that the selective binding agents may also bind orthologs of the polypeptide as set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6, that is, interspecies versions thereof, such as mouse and rat GPCR polypeptides.

The term "transduction" is used to refer to the transfer of genes from one bacterium to another, usually by a phage. "Transduction" also refers to the acquisition and transfer of eukaryotic cellular sequences by retroviruses.

The term "transfection" is used to refer to the uptake of foreign or exogenous DNA by a cell, and a cell has been "transfected" when the exogenous DNA has been introduced inside the cell membrane. A number of transfection techniques are well known in the art and are disclosed herein. See, e.g., Graham *et al.*, 1973, *Virology* 52:456; Sambrook *et al.*, *Molecular Cloning, A Laboratory Manual* (Cold Spring Harbor Laboratories, 1989); Davis *et al.*, *Basic Methods in Molecular Biology* (Elsevier, 1986); and Chu *et al.*, 1981, *Gene* 13:197. Such techniques can be used to introduce one or more exogenous DNA moieties into suitable host cells.

The term "transformation" as used herein refers to a change in a cell's genetic characteristics, and a cell has been transformed when it has been modified to contain a new DNA. For example, a cell is transformed where it is genetically modified from its native state. Following transfection or transduction, the transforming DNA may recombine with that of the cell by physically integrating into a chromosome of the cell, may be maintained transiently as an episomal element without being replicated, or may replicate independently as a plasmid. A cell is considered to have been stably transformed when the DNA is replicated with the division of the cell.

#### Relatedness of Nucleic Acid Molecules and/or Polypeptides

It is understood that related nucleic acid molecules include allelic or splice variants of the nucleic acid molecule of any of SEQ ID NO: 1, SEQ ID NO: 3, or SEQ ID NO: 5, and include sequences which are complementary to any of the above nucleotide sequences. Related nucleic acid molecules also include a nucleotide sequence encoding a polypeptide comprising or consisting essentially of a substitution, modification, addition and/or deletion of one or more amino acid residues compared to the polypeptide as set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6. Such related GPCR polypeptides may comprise, for example, an addition and/or a deletion of one or more N-linked or O-linked glycosylation sites or an addition and/or a deletion of one or more cysteine residues.

Related nucleic acid molecules also include fragments of GPCR nucleic acid molecules which encode a polypeptide of at least about 25 contiguous amino acids, or about 50 amino acids, or about 75 amino acids, or about 100 amino acids, or about 150 amino acids, or about 200 amino acids, or more than 200 amino acid residues of the GPCR polypeptide of any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6.

In addition, related GPCR nucleic acid molecules also include those molecules which comprise nucleotide sequences which hybridize under moderately or highly stringent conditions as defined herein with the fully complementary sequence of the GPCR nucleic acid molecule of any of SEQ ID NO: 1, SEQ ID NO: 3, or SEQ ID NO: 5, or of a molecule encoding a polypeptide, which polypeptide comprises the amino acid sequence as shown in any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6, or of a nucleic acid fragment as defined herein, or of a nucleic acid fragment encoding a polypeptide as defined herein. Hybridization probes may be prepared using the GPCR sequences provided herein to screen cDNA, genomic or synthetic DNA libraries for related sequences. Regions of the DNA and/or amino acid sequence of GPCR polypeptide that exhibit significant identity to known sequences are readily determined using sequence alignment algorithms as described herein and those regions may be used to design probes for screening.

The term "highly stringent conditions" refers to those conditions that are designed to permit hybridization of DNA strands whose sequences are highly complementary, and to exclude hybridization of significantly mismatched DNAs. Hybridization stringency is principally determined by temperature, ionic strength, and the concentration of denaturing agents such as formamide. Examples of "highly stringent conditions" for hybridization and washing are 0.015 M sodium chloride, 0.0015 M sodium citrate at 65-68°C or 0.015 M sodium chloride, 0.0015 M sodium citrate, and 50% formamide at 42°C. See Sambrook, Fritsch & Maniatis, *Molecular Cloning: A Laboratory Manual* (2nd ed., Cold Spring Harbor Laboratory, 1989); Anderson *et al.*, *Nucleic Acid Hybridisation: A Practical Approach* Ch. 4 (IRL Press Limited).

More stringent conditions (such as higher temperature, lower ionic strength, higher formamide, or other denaturing agent) may also be used – however, the rate of hybridization will be affected. Other agents may be included in the hybridization and washing buffers for the purpose of reducing non-specific and/or background hybridization. Examples are 0.1% bovine serum albumin, 0.1% polyvinylpyrrolidone, 0.1% sodium pyrophosphate, 0.1% sodium dodecylsulfate, NaDodSO<sub>4</sub>, (SDS), ficoll, Denhardt's solution, sonicated salmon sperm DNA (or another non-complementary DNA), and dextran sulfate, although other suitable agents can also be used. The concentration and types of these additives can be changed without substantially affecting the stringency of the hybridization conditions. Hybridization

experiments are usually carried out at pH 6.8-7.4; however, at typical ionic strength conditions, the rate of hybridization is nearly independent of pH. See Anderson *et al.*, *Nucleic Acid Hybridisation: A Practical Approach* Ch. 4 (IRL Press Limited).

Factors affecting the stability of DNA duplex include base composition, length, and degree of base pair mismatch. Hybridization conditions can be adjusted by one skilled in the art in order to accommodate these variables and allow DNAs of different sequence relatedness to form hybrids. The melting temperature of a perfectly matched DNA duplex can be estimated by the following equation:

$$T_m(^{\circ}\text{C}) = 81.5 + 16.6(\log[\text{Na}^+]) + 0.41(\% \text{G+C}) - 600/N - 0.72(\% \text{formamide})$$

where N is the length of the duplex formed,  $[\text{Na}^+]$  is the molar concentration of the sodium ion in the hybridization or washing solution, %G+C is the percentage of (guanine+cytosine) bases in the hybrid. For imperfectly matched hybrids, the melting temperature is reduced by approximately 1°C for each 1% mismatch.

The term "moderately stringent conditions" refers to conditions under which a DNA duplex with a greater degree of base pair mismatching than could occur under "highly stringent conditions" is able to form. Examples of typical "moderately stringent conditions" are 0.015 M sodium chloride, 0.0015 M sodium citrate at 50-65°C or 0.015 M sodium chloride, 0.0015 M sodium citrate, and 20% formamide at 37-50°C. By way of example, "moderately stringent conditions" of 50°C in 0.015 M sodium ion will allow about a 21% mismatch.

It will be appreciated by those skilled in the art that there is no absolute distinction between "highly stringent conditions" and "moderately stringent conditions." For example, at 0.015 M sodium ion (no formamide), the melting temperature of perfectly matched long DNA is about 71°C. With a wash at 65°C (at the same ionic strength), this would allow for approximately a 6% mismatch. To capture more distantly related sequences, one skilled in the art can simply lower the temperature or raise the ionic strength.

A good estimate of the melting temperature in 1M NaCl\* for oligonucleotide probes up to about 20nt is given by:

$$T_m = 2^{\circ}\text{C per A-T base pair} + 4^{\circ}\text{C per G-C base pair}$$

\*The sodium ion concentration in 6X salt sodium citrate (SSC) is 1M. See Suggs *et al.*, *Developmental Biology Using Purified Genes* 683 (Brown and Fox, eds., 1981).

High stringency washing conditions for oligonucleotides are usually at a temperature of 0-5°C below the  $T_m$  of the oligonucleotide in 6X SSC, 0.1% SDS.

In another embodiment, related nucleic acid molecules comprise or consist of a nucleotide sequence that is at least about 70 percent identical to the nucleotide sequence as shown in any of SEQ ID NO: 1, SEQ ID NO: 3, or SEQ ID NO: 5. In preferred embodiments, the nucleotide sequences are about 75 percent, or about 80 percent, or about 85 percent, or about 90 percent, or about 95, 96, 97, 98, or 99 percent identical to the nucleotide sequence as shown in any of SEQ ID NO: 1, SEQ ID NO: 3, or SEQ ID NO: 5. Related nucleic acid molecules encode polypeptides possessing at least one activity of the polypeptide set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6.

10 Differences in the nucleic acid sequence may result in conservative and/or non-conservative modifications of the amino acid sequence relative to the amino acid sequence of any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6.

Conservative modifications to the amino acid sequence of any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6 (and the corresponding modifications to the encoding nucleotides) will produce a polypeptide having functional and chemical characteristics similar to those of GPCR polypeptides. In contrast, substantial modifications in the functional and/or chemical characteristics of GPCR polypeptides may be accomplished by selecting substitutions in the amino acid sequence of any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6 that differ significantly in their effect on maintaining (a) the structure of the molecular backbone in the area of the substitution, for example, as a sheet or helical conformation, (b) the charge or hydrophobicity of the molecule at the target site, or (c) the bulk of the side chain.

For example, a "conservative amino acid substitution" may involve a substitution of a native amino acid residue with a nonnative residue such that there is little or no effect on the polarity or charge of the amino acid residue at that position. Furthermore, any native residue in the polypeptide may also be substituted with alanine, as has been previously described for "alanine scanning mutagenesis."

Conservative amino acid substitutions also encompass non-naturally occurring amino acid residues that are typically incorporated by chemical peptide synthesis rather than by synthesis in biological systems. These include peptidomimetics, and other reversed or inverted forms of amino acid moieties.

Naturally occurring residues may be divided into classes based on common side chain properties:

1) hydrophobic: norleucine, Met, Ala, Val, Leu, Ile;

- 2) neutral hydrophilic: Cys, Ser, Thr;
- 3) acidic: Asp, Glu;
- 4) basic: Asn, Gln, His, Lys, Arg;
- 5) residues that influence chain orientation: Gly, Pro; and
- 6) aromatic: Trp, Tyr, Phe.

For example, non-conservative substitutions may involve the exchange of a member of one of these classes for a member from another class. Such substituted residues may be introduced into regions of the human GPCR polypeptide that are homologous with non-human GPCR polypeptides, or into the non-homologous regions of the molecule.

In making such changes, the hydropathic index of amino acids may be considered. Each amino acid has been assigned a hydropathic index on the basis of its hydrophobicity and charge characteristics. The hydropathic indices are: isoleucine (+4.5); valine (+4.2); leucine (+3.8); phenylalanine (+2.8); cysteine/cystine (+2.5); methionine (+1.9); alanine (+1.8); glycine (-0.4); threonine (-0.7); serine (-0.8); tryptophan (-0.9); tyrosine (-1.3); proline (-1.6); histidine (-3.2); glutamate (-3.5); glutamine (-3.5); aspartate (-3.5); asparagine (-3.5); lysine (-3.9); and arginine (-4.5).

The importance of the hydropathic amino acid index in conferring interactive biological function on a protein is generally understood in the art (Kyte *et al.*, 1982, *J. Mol. Biol.* 157:105-31). It is known that certain amino acids may be substituted for other amino acids having a similar hydropathic index or score and still retain a similar biological activity. In making changes based upon the hydropathic index, the substitution of amino acids whose hydropathic indices are within  $\pm 2$  is preferred, those that are within  $\pm 1$  are particularly preferred, and those within  $\pm 0.5$  are even more particularly preferred.

It is also understood in the art that the substitution of like amino acids can be made effectively on the basis of hydrophilicity, particularly where the biologically functionally equivalent protein or peptide thereby created is intended for use in immunological embodiments, as in the present case. The greatest local average hydrophilicity of a protein, as governed by the hydrophilicity of its adjacent amino acids, correlates with its immunogenicity and antigenicity, *i.e.*, with a biological property of the protein.

The following hydrophilicity values have been assigned to these amino acid residues: arginine (+3.0); lysine (+3.0); aspartate (+3.0  $\pm$  1); glutamate (+3.0  $\pm$  1);



serine (+0.3); asparagine (+0.2); glutamine (+0.2); glycine (0); threonine (-0.4); proline (-0.5  $\pm$  1); alanine (-0.5); histidine (-0.5); cysteine (-1.0); methionine (-1.3); valine (-1.5); leucine (-1.8); isoleucine (-1.8); tyrosine (-2.3); phenylalanine (-2.5); and tryptophan (-3.4). In making changes based upon similar hydrophilicity values, the substitution of amino acids whose hydrophilicity values are within  $\pm 2$  is preferred, those that are within  $\pm 1$  are particularly preferred, and those within  $\pm 0.5$  are even more particularly preferred. One may also identify epitopes from primary amino acid sequences on the basis of hydrophilicity. These regions are also referred to as "epitopic core regions."

Desired amino acid substitutions (whether conservative or non-conservative) can be determined by those skilled in the art at the time such substitutions are desired. For example, amino acid substitutions can be used to identify important residues of the GPCR polypeptide, or to increase or decrease the affinity of the GPCR polypeptides described herein. Exemplary amino acid substitutions are set forth in Table I.

Table I  
Amino Acid Substitutions

Original Residues	Exemplary Substitutions	Preferred Substitutions
Ala	Val, Leu, Ile	Val
Arg	Lys, Gln, Asn	Lys
Asn	Gln	Gln
Asp	Glu	Glu
Cys	Ser, Ala	Ser
Gln	Asn	Asn
Glu	Asp	Asp
Gly	Pro, Ala	Ala
His	Asn, Gln, Lys, Arg	Arg
Ile	Leu, Val, Met, Ala, Phe, Norleucine	Leu
Leu	Norleucine, Ile, Val, Met, Ala, Phe	Ile

Lys	Arg, 1,4 Diamino-butyric Acid, Gln, Asn	Arg
Met	Leu, Phe, Ile	Leu
Phe	Leu, Val, Ile, Ala, Tyr	Leu
Pro	Ala	Gly
Ser	Thr, Ala, Cys	Thr
Thr	Ser	Ser
Trp	Tyr, Phe	Tyr
Tyr	Trp, Phe, Thr, Ser	Phe
Val	Ile, Met, Leu, Phe, Ala, Norleucine	Leu

A skilled artisan will be able to determine suitable variants of the polypeptide as set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6 using well-known techniques. For identifying suitable areas of the molecule that may be changed without destroying biological activity, one skilled in the art may target areas not believed to be important for activity. For example, when similar polypeptides with similar activities from the same species or from other species are known, one skilled in the art may compare the amino acid sequence of a GPCR polypeptide to such similar polypeptides. With such a comparison, one can identify residues and portions of the molecules that are conserved among similar polypeptides. It will be appreciated that changes in areas of the GPCR molecule that are not conserved relative to such similar polypeptides would be less likely to adversely affect the biological activity and/or structure of a GPCR polypeptide. One skilled in the art would also know that, even in relatively conserved regions, one may substitute chemically similar amino acids for the naturally occurring residues while retaining activity (conservative amino acid residue substitutions). Therefore, even areas that may be important for biological activity or for structure may be subject to conservative amino acid substitutions without destroying the biological activity or without adversely affecting the polypeptide structure.

Additionally, one skilled in the art can review structure-function studies identifying residues in similar polypeptides that are important for activity or structure.

In view of such a comparison, one can predict the importance of amino acid residues in a GPCR polypeptide that correspond to amino acid residues that are important for activity or structure in similar polypeptides. One skilled in the art may opt for chemically similar amino acid substitutions for such predicted important amino acid residues of GPCR polypeptides.

One skilled in the art can also analyze the three-dimensional structure and amino acid sequence in relation to that structure in similar polypeptides. In view of such information, one skilled in the art may predict the alignment of amino acid residues of GPCR polypeptide with respect to its three dimensional structure. One skilled in the art may choose not to make radical changes to amino acid residues predicted to be on the surface of the protein, since such residues may be involved in important interactions with other molecules. Moreover, one skilled in the art may generate test variants containing a single amino acid substitution at each amino acid residue. The variants could be screened using activity assays known to those with skill in the art. Such variants could be used to gather information about suitable variants. For example, if one discovered that a change to a particular amino acid residue resulted in destroyed, undesirably reduced, or unsuitable activity, variants with such a change would be avoided. In other words, based on information gathered from such routine experiments, one skilled in the art can readily determine the amino acids where further substitutions should be avoided either alone or in combination with other mutations.

A number of scientific publications have been devoted to the prediction of secondary structure. See Moulton, 1996, *Curr. Opin. Biotechnol.* 7:422-27; Chou *et al.*, 1974, *Biochemistry* 13:222-45; Chou *et al.*, 1974, *Biochemistry* 113:211-22; Chou *et al.*, 1978, *Adv. Enzymol. Relat. Areas Mol. Biol.* 47:45-48; Chou *et al.*, 1978, *Ann. Rev. Biochem.* 47:251-276; and Chou *et al.*, 1979, *Biophys. J.* 26:367-84. Moreover, computer programs are currently available to assist with predicting secondary structure. One method of predicting secondary structure is based upon homology modeling. For example, two polypeptides or proteins that have a sequence identity of greater than 30%, or similarity greater than 40%, often have similar structural topologies. The recent growth of the protein structural database (PDB) has provided enhanced predictability of secondary structure, including the potential number of folds within the structure of a polypeptide or protein. See Holm *et al.*, 1999, *Nucleic Acids Res.* 27:244-47. It has been suggested that there are a limited number of folds

in a given polypeptide or protein and that once a critical number of structures have been resolved, structural prediction will become dramatically more accurate (Brenner *et al.*, 1997, *Curr. Opin. Struct. Biol.* 7:369-76).

Additional methods of predicting secondary structure include "threading" (Jones, 1997, *Curr. Opin. Struct. Biol.* 7:377-87; Sippl *et al.*, 1996, *Structure* 4:15-19), "profile analysis" (Bowie *et al.*, 1991, *Science*, 253:164-70; Gribskov *et al.*, 1990, *Methods Enzymol.* 183:146-59; Gribskov *et al.*, 1987, *Proc. Nat. Acad. Sci. U.S.A.* 84:4355-58), and "evolutionary linkage" (See Holm *et al.*, *supra*, and Brenner *et al.*, *supra*).

Preferred GPCR polypeptide variants include glycosylation variants wherein the number and/or type of glycosylation sites have been altered compared to the amino acid sequence set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6. In one embodiment, GPCR polypeptide variants comprise a greater or a lesser number of N-linked glycosylation sites than the amino acid sequence set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6. An N-linked glycosylation site is characterized by the sequence: Asn-X-Ser or Asn-X-Thr, wherein the amino acid residue designated as X may be any amino acid residue except proline. The substitution of amino acid residues to create this sequence provides a potential new site for the addition of an N-linked carbohydrate chain. Alternatively, substitutions that eliminate this sequence will remove an existing N-linked carbohydrate chain. Also provided is a rearrangement of N-linked carbohydrate chains wherein one or more N-linked glycosylation sites (typically those that are naturally occurring) are eliminated and one or more new N-linked sites are created. Additional preferred GPCR variants include cysteine variants, wherein one or more cysteine residues are deleted or substituted with another amino acid (*e.g.*, serine) as compared to the amino acid sequence set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6. Cysteine variants are useful when GPCR polypeptides must be refolded into a biologically active conformation such as after the isolation of insoluble inclusion bodies. Cysteine variants generally have fewer cysteine residues than the native protein, and typically have an even number to minimize interactions resulting from unpaired cysteines.

In other embodiments, GPCR polypeptide variants comprise an amino acid sequence as set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6 with at least one amino acid insertion and wherein the polypeptide has an activity of the

polypeptide set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6, or an amino acid sequence as set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6 with at least one amino acid deletion and wherein the polypeptide has an activity of the polypeptide set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6. GPCR polypeptide variants also comprise an amino acid sequence as set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6 wherein the polypeptide has a carboxyl- and/or amino-terminal truncation and further wherein the polypeptide has an activity of the polypeptide set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6. GPCR polypeptide variants further comprise an amino acid sequence as set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6 with at least one modification that is an amino acid substitution, an amino acid insertion, an amino acid deletion, carboxyl-terminal truncation, or amino-terminal truncation and wherein the polypeptide has an activity of the polypeptide set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6.

In further embodiments, GPCR polypeptide variants comprise an amino acid sequence that is at least about 70 percent identical to the amino acid sequence as set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6. In preferred embodiments, GPCR polypeptide variants comprise an amino acid sequence that is at least about 75 percent, or about 80 percent, or about 85 percent, or about 90 percent, or about 95, 96, 97, 98, or 99 percent identical percent to the amino acid sequence as set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6. GPCR polypeptide variants possess at least one activity of the polypeptide set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6.

In addition, the polypeptide comprising the amino acid sequence of any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6, or other GPCR polypeptide, may be fused to a homologous polypeptide to form a homodimer or to a heterologous polypeptide to form a heterodimer. Heterologous peptides and polypeptides include, but are not limited to: an epitope to allow for the detection and/or isolation of a GPCR fusion polypeptide; a transmembrane receptor protein or a portion thereof, such as an extracellular domain or a transmembrane and intracellular domain; a ligand or a portion thereof which binds to a transmembrane receptor protein; an enzyme or portion thereof which is catalytically active; a polypeptide or peptide which promotes oligomerization, such as a leucine zipper domain; a polypeptide or peptide which increases stability, such as an immunoglobulin constant region; and a polypeptide

which has a therapeutic activity different from the polypeptide comprising the amino acid sequence as set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6, or other GPCR polypeptide.

Fusions can be made either at the amino-terminus or at the carboxyl-terminus of the polypeptide comprising the amino acid sequence set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6, or other GPCR polypeptide. Fusions may be direct with no linker or adapter molecule or may be through a linker or adapter molecule. A linker or adapter molecule may be one or more amino acid residues, typically from about 20 to about 50 amino acid residues. A linker or adapter molecule may also be designed with a cleavage site for a DNA restriction endonuclease or for a protease to allow for the separation of the fused moieties. It will be appreciated that once constructed, the fusion polypeptides can be derivatized according to the methods described herein.

In a further embodiment of the invention, the polypeptide comprising the amino acid sequence of any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6, or other GPCR polypeptide, is fused to one or more domains of an Fc region of human IgG. Antibodies comprise two functionally independent parts, a variable domain known as "Fab," that binds an antigen, and a constant domain known as "Fc," that is involved in effector functions such as complement activation and attack by phagocytic cells. An Fc has a long serum half-life, whereas an Fab is short-lived. Capon *et al.*, 1989, *Nature* 337:525-31. When constructed together with a therapeutic protein, an Fc domain can provide longer half-life or incorporate such functions as Fc receptor binding, protein A binding, complement fixation, and perhaps even placental transfer. *Id.* Table II summarizes the use of certain Fc fusions known in the art.

Table II

Fc Fusion with Therapeutic Proteins

Form of Fc	Fusion partner	Therapeutic implications	Reference
IgG1	N-terminus of CD30-L	Hodgkin's disease; anaplastic lymphoma; T-cell leukemia	U.S. Patent No. 5,480,981
Murine Fcγ2a	IL-10	anti-inflammatory; transplant rejection	Zheng <i>et al.</i> , 1995, <i>J. Immunol.</i> 154:5590-600
IgG1	TNF receptor	septic shock	Fisher <i>et al.</i> , 1996, <i>N. Engl. J. Med.</i> 334:1697-1702; Van Zee <i>et al.</i> ,

			1996, <i>J. Immunol.</i> 156:2221-30
IgG, IgA, IgM, or IgE (excluding the first domain)	TNF receptor	inflammation, autoimmune disorders	U.S. Patent No. 5,808,029
IgG1	CD4 receptor	AIDS	Capon <i>et al.</i> , 1989, <i>Nature</i> 337: 525-31
IgG1, IgG3	N-terminus of IL-2	anti-cancer, antiviral	Harvill <i>et al.</i> , 1995, <i>Immunotech.</i> 1:95-105
IgG1	C-terminus of OPG	osteoarthritis; bone density	WO 97/23614
IgG1	N-terminus of leptin	anti-obesity	PCT/US 97/23183, filed December 11, 1997
Human Ig C $\gamma$ 1	CTLA-4	autoimmune disorders	Linsley, 1991, <i>J. Exp.</i> <i>Med.</i> , 174:561-69

In one example, a human IgG hinge, CH2, and CH3 region may be fused at either the amino-terminus or carboxyl-terminus of the GPCR polypeptides using methods known to the skilled artisan. In another example, a human IgG hinge, CH2, and CH3 region may be fused at either the amino-terminus or carboxyl-terminus of a GPCR polypeptide fragment (*e.g.*, the predicted extracellular portion of GPCR polypeptide).

The resulting GPCR fusion polypeptide may be purified by use of a Protein A affinity column. Peptides and proteins fused to an Fc region have been found to exhibit a substantially greater half-life *in vivo* than the unfused counterpart. Also, a fusion to an Fc region allows for dimerization/multimerization of the fusion polypeptide. The Fc region may be a naturally occurring Fc region, or may be altered to improve certain qualities, such as therapeutic qualities, circulation time, or reduced aggregation.

Identity and similarity of related nucleic acid molecules and polypeptides are readily calculated by known methods. Such methods include, but are not limited to those described in *Computational Molecular Biology* (A.M. Lesk, ed., Oxford University Press 1988); *Biocomputing: Informatics and Genome Projects* (D.W. Smith, ed., Academic Press 1993); *Computer Analysis of Sequence Data* (Part 1, A.M. Griffin and H.G. Griffin, eds., Humana Press 1994); G. von Heinle, *Sequence Analysis in Molecular Biology* (Academic Press 1987); *Sequence Analysis Primer* (M. Gribskov and J. Devereux, eds., M. Stockton Press 1991); and Carillo *et al.*, 1988, *SIAM J. Applied Math.*, 48:1073.

Preferred methods to determine identity and/or similarity are designed to give the largest match between the sequences tested. Methods to determine identity and similarity are described in publicly available computer programs. Preferred computer program methods to determine identity and similarity between two sequences include, 5 but are not limited to, the GCG program package, including GAP (Devereux *et al.*, 1984, *Nucleic Acids Res.* 12:387; Genetics Computer Group, University of Wisconsin, Madison, WI), BLASTP, BLASTN, and FASTA (Altschul *et al.*, 1990, *J. Mol. Biol.* 215:403-10). The BLASTX program is publicly available from the National Center for Biotechnology Information (NCBI) and other sources (Altschul *et al.*, 10 *BLAST Manual* (NCB NLM NIH, Bethesda, MD); Altschul *et al.*, 1990, *supra*). The well-known Smith Waterman algorithm may also be used to determine identity.

Certain alignment schemes for aligning two amino acid sequences may result in the matching of only a short region of the two sequences, and this small aligned region may have very high sequence identity even though there is no significant 15 relationship between the two full-length sequences. Accordingly, in a preferred embodiment, the selected alignment method (GAP program) will result in an alignment that spans at least 50 contiguous amino acids of the claimed polypeptide.

For example, using the computer algorithm GAP (Genetics Computer Group, University of Wisconsin, Madison, WI), two polypeptides for which the percent 20 sequence identity is to be determined are aligned for optimal matching of their respective amino acids (the "matched span," as determined by the algorithm). A gap opening penalty (which is calculated as 3X the average diagonal; the "average diagonal" is the average of the diagonal of the comparison matrix being used; the "diagonal" is the score or number assigned to each perfect amino acid match by the 25 particular comparison matrix) and a gap extension penalty (which is usually 0.1X the gap opening penalty), as well as a comparison matrix such as PAM 250 or BLOSUM 62 are used in conjunction with the algorithm. A standard comparison matrix is also used by the algorithm (*see Dayhoff et al.*, 5 *Atlas of Protein Sequence and Structure* (Supp. 3 1978)(PAM250 comparison matrix); Henikoff *et al.*, 1992, *Proc. Natl. Acad. Sci USA* 89:10915-19 (BLOSUM 62 comparison matrix)). 30

Preferred parameters for polypeptide sequence comparison include the following:

Algorithm: Needleman and Wunsch, 1970, *J. Mol. Biol.* 48:443-53;



Comparison matrix: BLOSUM 62 (Henikoff *et al.*, *supra*);

Gap Penalty: 12

Gap Length Penalty: 4

Threshold of Similarity: 0

5

The GAP program is useful with the above parameters. The aforementioned parameters are the default parameters for polypeptide comparisons (along with no penalty for end gaps) using the GAP algorithm.

Preferred parameters for nucleic acid molecule sequence comparison include  
10 the following:

Algorithm: Needleman and Wunsch, *supra*;

Comparison matrix: matches = +10, mismatch = 0

Gap Penalty: 50

15 Gap Length Penalty: 3

The GAP program is also useful with the above parameters. The aforementioned parameters are the default parameters for nucleic acid molecule comparisons.

Other exemplary algorithms, gap opening penalties, gap extension penalties,  
20 comparison matrices, and thresholds of similarity may be used, including those set forth in the Program Manual, Wisconsin Package, Version 9, September, 1997. The particular choices to be made will be apparent to those of skill in the art and will depend on the specific comparison to be made, such as DNA-to-DNA, protein-to-protein, protein-to-DNA; and additionally, whether the comparison is between given  
25 pairs of sequences (in which case GAP or BestFit are generally preferred) or between one sequence and a large database of sequences (in which case FASTA or BLASTA are preferred).

#### Nucleic Acid Molecules

30 The nucleic acid molecules encoding a polypeptide comprising the amino acid sequence of a GPCR polypeptide can readily be obtained in a variety of ways including, without limitation, chemical synthesis, cDNA or genomic library screening, expression library screening, and/or PCR amplification of cDNA.

Recombinant DNA methods used herein are generally those set forth in Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, 1989) and/or *Current Protocols in Molecular Biology* (Ausubel *et al.*, eds., Green Publishers Inc. and Wiley and Sons 1994). The invention provides for  
5 nucleic acid molecules as described herein and methods for obtaining such molecules.

Where a gene encoding the amino acid sequence of a GPCR polypeptide has been identified from one species, all or a portion of that gene may be used as a probe to identify orthologs or related genes from the same species. The probes or primers may be used to screen cDNA libraries from various tissue sources believed to express  
10 the GPCR polypeptide. In addition, part or all of a nucleic acid molecule having the sequence as set forth in any of SEQ ID NO: 1, SEQ ID NO: 3, or SEQ ID NO: 5 may be used to screen a genomic library to identify and isolate a gene encoding the amino acid sequence of a GPCR polypeptide. Typically, conditions of moderate or high stringency will be employed for screening to minimize the number of false positives  
15 obtained from the screening.

Nucleic acid molecules encoding the amino acid sequence of GPCR polypeptides may also be identified by expression cloning which employs the detection of positive clones based upon a property of the expressed protein. Typically, nucleic acid libraries are screened by the binding an antibody or other  
20 binding partner (*e.g.*, receptor or ligand) to cloned proteins that are expressed and displayed on a host cell surface. The antibody or binding partner is modified with a detectable label to identify those cells expressing the desired clone.

Recombinant expression techniques conducted in accordance with the descriptions set forth below may be followed to produce these polynucleotides and to  
25 express the encoded polypeptides. For example, by inserting a nucleic acid sequence that encodes the amino acid sequence of a GPCR polypeptide into an appropriate vector, one skilled in the art can readily produce large quantities of the desired nucleotide sequence. The sequences can then be used to generate detection probes or amplification primers. Alternatively, a polynucleotide encoding the amino acid  
30 sequence of a GPCR polypeptide can be inserted into an expression vector. By introducing the expression vector into an appropriate host, the encoded GPCR polypeptide may be produced in large amounts.

Another method for obtaining a suitable nucleic acid sequence is the polymerase chain reaction (PCR). In this method, cDNA is prepared from

poly(A)+RNA or total RNA using the enzyme reverse transcriptase. Two primers, typically complementary to two separate regions of cDNA encoding the amino acid sequence of a GPCR polypeptide, are then added to the cDNA along with a polymerase such as *Taq* polymerase, and the polymerase amplifies the cDNA region  
5 between the two primers.

Another means of preparing a nucleic acid molecule encoding the amino acid sequence of a GPCR polypeptide is chemical synthesis using methods well known to the skilled artisan such as those described by Engels *et al.*, 1989, *Angew. Chem. Intl. Ed.* 28:716-34. These methods include, *inter alia*, the phosphotriester,  
10 phosphoramidite, and H-phosphonate methods for nucleic acid synthesis. A preferred method for such chemical synthesis is polymer-supported synthesis using standard phosphoramidite chemistry. Typically, the DNA encoding the amino acid sequence of a GPCR polypeptide will be several hundred nucleotides in length. Nucleic acids larger than about 100 nucleotides can be synthesized as several fragments using these  
15 methods. The fragments can then be ligated together to form the full-length nucleotide sequence of a GPCR gene. Usually, the DNA fragment encoding the amino-terminus of the polypeptide will have an ATG, which encodes a methionine residue. This methionine may or may not be present on the mature form of the GPCR polypeptide, depending on whether the polypeptide produced in the host cell is  
20 designed to be secreted from that cell. Other methods known to the skilled artisan may be used as well.

In certain embodiments, nucleic acid variants contain codons which have been altered for optimal expression of a GPCR polypeptide in a given host cell. Particular codon alterations will depend upon the GPCR polypeptide and host cell selected for  
25 expression. Such "codon optimization" can be carried out by a variety of methods, for example, by selecting codons which are preferred for use in highly expressed genes in a given host cell. Computer algorithms which incorporate codon frequency tables such as "Eco\_high.Cod" for codon preference of highly expressed bacterial genes may be used and are provided by the University of Wisconsin Package Version  
30 9.0 (Genetics Computer Group, Madison, WI). Other useful codon frequency tables include "Celegans\_high.cod," "Celegans\_low.cod," "Drosophila\_high.cod," "Human\_high.cod," "Maize\_high.cod," and "Yeast\_high.cod."

In some cases, it may be desirable to prepare nucleic acid molecules encoding GPCR polypeptide variants. Nucleic acid molecules encoding variants may be

produced using site directed mutagenesis, PCR amplification, or other appropriate methods, where the primer(s) have the desired point mutations (*see* Sambrook *et al.*, *supra*, and Ausubel *et al.*, *supra*, for descriptions of mutagenesis techniques). Chemical synthesis using methods described by Engels *et al.*, *supra*, may also be used  
5 to prepare such variants. Other methods known to the skilled artisan may be used as well.

### Vectors and Host Cells

A nucleic acid molecule encoding the amino acid sequence of a GPCR  
10 polypeptide is inserted into an appropriate expression vector using standard ligation techniques. The vector is typically selected to be functional in the particular host cell employed (*i.e.*, the vector is compatible with the host cell machinery such that amplification of the gene and/or expression of the gene can occur). A nucleic acid molecule encoding the amino acid sequence of a GPCR polypeptide may be  
15 amplified/expressed in prokaryotic, yeast, insect (baculovirus systems) and/or eukaryotic host cells. Selection of the host cell will depend in part on whether a GPCR polypeptide is to be post-translationally modified (*e.g.*, glycosylated and/or phosphorylated). If so, yeast, insect, or mammalian host cells are preferable. For a review of expression vectors, *see Meth. Enz.*, vol. 185 (D.V. Goeddel, ed., Academic  
20 Press 1990).

Typically, expression vectors used in any of the host cells will contain sequences for plasmid maintenance and for cloning and expression of exogenous nucleotide sequences. Such sequences, collectively referred to as “flanking sequences” in certain embodiments will typically include one or more of the  
25 following nucleotide sequences: a promoter, one or more enhancer sequences, an origin of replication, a transcriptional termination sequence, a complete intron sequence containing a donor and acceptor splice site, a sequence encoding a leader sequence for polypeptide secretion, a ribosome binding site, a polyadenylation sequence, a polylinker region for inserting the nucleic acid encoding the polypeptide  
30 to be expressed, and a selectable marker element. Each of these sequences is discussed below.

Optionally, the vector may contain a “tag”-encoding sequence, *i.e.*, an oligonucleotide molecule located at the 5' or 3' end of the GPCR polypeptide coding sequence; the oligonucleotide sequence encodes polyHis (such as hexaHis), or

another "tag" such as FLAG, HA (hemagglutinin influenza virus), or *myc* for which commercially available antibodies exist. This tag is typically fused to the polypeptide upon expression of the polypeptide, and can serve as a means for affinity purification of the GPCR polypeptide from the host cell. Affinity purification can be accomplished, for example, by column chromatography using antibodies against the tag as an affinity matrix. Optionally, the tag can subsequently be removed from the purified GPCR polypeptide by various means such as using certain peptidases for cleavage.

Flanking sequences may be homologous (*i.e.*, from the same species and/or strain as the host cell), heterologous (*i.e.*, from a species other than the host cell species or strain), hybrid (*i.e.*, a combination of flanking sequences from more than one source), or synthetic, or the flanking sequences may be native sequences that normally function to regulate GPCR polypeptide expression. As such, the source of a flanking sequence may be any prokaryotic or eukaryotic organism, any vertebrate or invertebrate organism, or any plant, provided that the flanking sequence is functional in, and can be activated by, the host cell machinery.

Flanking sequences useful in the vectors of this invention may be obtained by any of several methods well known in the art. Typically, flanking sequences useful herein – other than the GPCR gene flanking sequences – will have been previously identified by mapping and/or by restriction endonuclease digestion and can thus be isolated from the proper tissue source using the appropriate restriction endonucleases. In some cases, the full nucleotide sequence of a flanking sequence may be known. Here, the flanking sequence may be synthesized using the methods described herein for nucleic acid synthesis or cloning.

Where all or only a portion of the flanking sequence is known, it may be obtained using PCR and/or by screening a genomic library with a suitable oligonucleotide and/or flanking sequence fragment from the same or another species. Where the flanking sequence is not known, a fragment of DNA containing a flanking sequence may be isolated from a larger piece of DNA that may contain, for example, a coding sequence or even another gene or genes. Isolation may be accomplished by restriction endonuclease digestion to produce the proper DNA fragment followed by isolation using agarose gel purification, Qiagen® column chromatography (Chatsworth, CA), or other methods known to the skilled artisan. The selection of

suitable enzymes to accomplish this purpose will be readily apparent to one of ordinary skill in the art.

An origin of replication is typically a part of those prokaryotic expression vectors purchased commercially, and the origin aids in the amplification of the vector in a host cell. Amplification of the vector to a certain copy number can, in some cases, be important for the optimal expression of a GPCR polypeptide. If the vector of choice does not contain an origin of replication site, one may be chemically synthesized based on a known sequence, and ligated into the vector. For example, the origin of replication from the plasmid pBR322 (New England Biolabs, Beverly, MA) is suitable for most gram-negative bacteria and various origins (*e.g.*, SV40, polyoma, adenovirus, vesicular stomatitis virus (VSV), or papillomaviruses such as HPV or BPV) are useful for cloning vectors in mammalian cells. Generally, the origin of replication component is not needed for mammalian expression vectors (for example, the SV40 origin is often used only because it contains the early promoter).

A transcription termination sequence is typically located 3' of the end of a polypeptide coding region and serves to terminate transcription. Usually, a transcription termination sequence in prokaryotic cells is a G-C rich fragment followed by a poly-T sequence. While the sequence is easily cloned from a library or even purchased commercially as part of a vector, it can also be readily synthesized using methods for nucleic acid synthesis such as those described herein.

A selectable marker gene element encodes a protein necessary for the survival and growth of a host cell grown in a selective culture medium. Typical selection marker genes encode proteins that (a) confer resistance to antibiotics or other toxins, *e.g.*, ampicillin, tetracycline, or kanamycin for prokaryotic host cells; (b) complement auxotrophic deficiencies of the cell; or (c) supply critical nutrients not available from complex media. Preferred selectable markers are the kanamycin resistance gene, the ampicillin resistance gene, and the tetracycline resistance gene. A neomycin resistance gene may also be used for selection in prokaryotic and eukaryotic host cells.

Other selection genes may be used to amplify the gene that will be expressed. Amplification is the process wherein genes that are in greater demand for the production of a protein critical for growth are reiterated in tandem within the chromosomes of successive generations of recombinant cells. Examples of suitable selectable markers for mammalian cells include dihydrofolate reductase (DHFR) and

thymidine kinase. The mammalian cell transformants are placed under selection pressure wherein only the transformants are uniquely adapted to survive by virtue of the selection gene present in the vector. Selection pressure is imposed by culturing the transformed cells under conditions in which the concentration of selection agent in the medium is successively changed, thereby leading to the amplification of both the selection gene and the DNA that encodes a GPCR polypeptide. As a result, increased quantities of GPCR polypeptide are synthesized from the amplified DNA.

A ribosome binding site is usually necessary for translation initiation of mRNA and is characterized by a Shine-Dalgarno sequence (prokaryotes) or a Kozak sequence (eukaryotes). The element is typically located 3' to the promoter and 5' to the coding sequence of a GPCR polypeptide to be expressed. The Shine-Dalgarno sequence is varied but is typically a polypurine (*i.e.*, having a high A-G content). Many Shine-Dalgarno sequences have been identified, each of which can be readily synthesized using methods set forth herein and used in a prokaryotic vector.

A leader, or signal, sequence may be used to direct a GPCR polypeptide out of the host cell. Typically, a nucleotide sequence encoding the signal sequence is positioned in the coding region of a GPCR nucleic acid molecule, or directly at the 5' end of a GPCR polypeptide coding region. Many signal sequences have been identified, and any of those that are functional in the selected host cell may be used in conjunction with a GPCR nucleic acid molecule. Therefore, a signal sequence may be homologous (naturally occurring) or heterologous to the GPCR nucleic acid molecule. Additionally, a signal sequence may be chemically synthesized using methods described herein. In most cases, the secretion of a GPCR polypeptide from the host cell via the presence of a signal peptide will result in the removal of the signal peptide from the secreted GPCR polypeptide. The signal sequence may be a component of the vector, or it may be a part of a GPCR nucleic acid molecule that is inserted into the vector.

Included within the scope of this invention is the use of either a nucleotide sequence encoding a native GPCR polypeptide signal sequence joined to a GPCR polypeptide coding region or a nucleotide sequence encoding a heterologous signal sequence joined to a GPCR polypeptide coding region. The heterologous signal sequence selected should be one that is recognized and processed, *i.e.*, cleaved by a signal peptidase, by the host cell. For prokaryotic host cells that do not recognize and process the native GPCR polypeptide signal sequence, the signal sequence is

substituted by a prokaryotic signal sequence selected, for example, from the group of the alkaline phosphatase, penicillinase, or heat-stable enterotoxin II leaders. For yeast secretion, the native GPCR polypeptide signal sequence may be substituted by the yeast invertase, alpha factor, or acid phosphatase leaders. In mammalian cell  
5 expression the native signal sequence is satisfactory, although other mammalian signal sequences may be suitable.

In some cases, such as where glycosylation is desired in a eukaryotic host cell expression system, one may manipulate the various presequences to improve glycosylation or yield. For example, one may alter the peptidase cleavage site of a  
10 particular signal peptide, or add pro-sequences, which also may affect glycosylation. The final protein product may have, in the -1 position (relative to the first amino acid of the mature protein) one or more additional amino acids incident to expression, which may not have been totally removed. For example, the final protein product may have one or two amino acid residues found in the peptidase cleavage site,  
15 attached to the amino-terminus. Alternatively, use of some enzyme cleavage sites may result in a slightly truncated form of the desired GPCR polypeptide, if the enzyme cuts at such area within the mature polypeptide.

In many cases, transcription of a nucleic acid molecule is increased by the presence of one or more introns in the vector; this is particularly true where a  
20 polypeptide is produced in eukaryotic host cells, especially mammalian host cells. The introns used may be naturally occurring within the GPCR gene especially where the gene used is a full-length genomic sequence or a fragment thereof. Where the intron is not naturally occurring within the gene (as for most cDNAs), the intron may be obtained from another source. The position of the intron with respect to flanking  
25 sequences and the GPCR gene is generally important, as the intron must be transcribed to be effective. Thus, when a GPCR cDNA molecule is being transcribed, the preferred position for the intron is 3' to the transcription start site and 5' to the poly-A transcription termination sequence. Preferably, the intron or introns will be located on one side or the other (*i.e.*, 5' or 3') of the cDNA such that it does not  
30 interrupt the coding sequence. Any intron from any source, including viral, prokaryotic and eukaryotic (plant or animal) organisms, may be used to practice this invention, provided that it is compatible with the host cell into which it is inserted. Also included herein are synthetic introns. Optionally, more than one intron may be used in the vector.



The expression and cloning vectors of the present invention will typically contain a promoter that is recognized by the host organism and operably linked to the molecule encoding the GPCR polypeptide. Promoters are untranscribed sequences located upstream (*i.e.*, 5') to the start codon of a structural gene (generally within  
5 about 100 to 1000 bp) that control the transcription of the structural gene. Promoters are conventionally grouped into one of two classes: inducible promoters and constitutive promoters. Inducible promoters initiate increased levels of transcription from DNA under their control in response to some change in culture conditions, such as the presence or absence of a nutrient or a change in temperature. Constitutive  
10 promoters, on the other hand, initiate continual gene product production; that is, there is little or no control over gene expression. A large number of promoters, recognized by a variety of potential host cells, are well known. A suitable promoter is operably linked to the DNA encoding GPCR polypeptide by removing the promoter from the source DNA by restriction enzyme digestion and inserting the desired promoter  
15 sequence into the vector. The native GPCR promoter sequence may be used to direct amplification and/or expression of a GPCR nucleic acid molecule. A heterologous promoter is preferred, however, if it permits greater transcription and higher yields of the expressed protein as compared to the native promoter, and if it is compatible with the host cell system that has been selected for use.

20 Promoters suitable for use with prokaryotic hosts include the beta-lactamase and lactose promoter systems; alkaline phosphatase; a tryptophan (*trp*) promoter system; and hybrid promoters such as the *tac* promoter. Other known bacterial promoters are also suitable. Their sequences have been published, thereby enabling one skilled in the art to ligate them to the desired DNA sequence, using linkers or  
25 adapters as needed to supply any useful restriction sites.

Suitable promoters for use with yeast hosts are also well known in the art. Yeast enhancers are advantageously used with yeast promoters. Suitable promoters for use with mammalian host cells are well known and include, but are not limited to, those obtained from the genomes of viruses such as polyoma virus, fowlpox virus,  
30 adenovirus (such as Adenovirus 2), bovine papilloma virus, avian sarcoma virus, cytomegalovirus, retroviruses, hepatitis-B virus and most preferably Simian Virus 40 (SV40). Other suitable mammalian promoters include heterologous mammalian promoters, for example, heat-shock promoters and the actin promoter.

Additional promoters which may be of interest in controlling GPCR gene expression include, but are not limited to: the SV40 early promoter region (Bernoist and Chambon, 1981, *Nature* 290:304-10); the CMV promoter; the promoter contained in the 3' long terminal repeat of Rous sarcoma virus (Yamamoto, *et al.*, 1980, *Cell* 5 22:787-97); the herpes thymidine kinase promoter (Wagner *et al.*, 1981, *Proc. Natl. Acad. Sci. U.S.A.* 78:1444-45); the regulatory sequences of the metallothioneine gene (Brinster *et al.*, 1982, *Nature* 296:39-42); prokaryotic expression vectors such as the beta-lactamase promoter (Villa-Kamaroff *et al.*, 1978, *Proc. Natl. Acad. Sci. U.S.A.*, 75:3727-31); or the tac promoter (DeBoer *et al.*, 1983, *Proc. Natl. Acad. Sci. U.S.A.*, 10 80:21-25). Also of interest are the following animal transcriptional control regions, which exhibit tissue specificity and have been utilized in transgenic animals: the elastase I gene control region which is active in pancreatic acinar cells (Swift *et al.*, 1984, *Cell* 38:639-46; Ornitz *et al.*, 1986, *Cold Spring Harbor Symp. Quant. Biol.* 50:399-409 (1986); MacDonald, 1987, *Hepatology* 7:425-515); the insulin gene 15 control region which is active in pancreatic beta cells (Hanahan, 1985, *Nature* 315:115-22); the immunoglobulin gene control region which is active in lymphoid cells (Grosschedl *et al.*, 1984, *Cell* 38:647-58; Adames *et al.*, 1985, *Nature* 318:533-38; Alexander *et al.*, 1987, *Mol. Cell. Biol.*, 7:1436-44); the mouse mammary tumor virus control region which is active in testicular, breast, lymphoid and mast cells 20 (Leder *et al.*, 1986, *Cell* 45:485-95); the albumin gene control region which is active in liver (Pinkert *et al.*, 1987, *Genes and Devel.* 1:268-76); the alpha-feto-protein gene control region which is active in liver (Krumlauf *et al.*, 1985, *Mol. Cell. Biol.*, 5:1639-48; Hammer *et al.*, 1987, *Science* 235:53-58); the alpha 1-antitrypsin gene control region which is active in the liver (Kelsey *et al.*, 1987, *Genes and Devel.* 1:161-71); 25 the beta-globin gene control region which is active in myeloid cells (Mogam *et al.*, 1985, *Nature* 315:338-40; Kollias *et al.*, 1986, *Cell* 46:89-94); the myelin basic protein gene control region which is active in oligodendrocyte cells in the brain (Readhead *et al.*, 1987, *Cell* 48:703-12); the myosin light chain-2 gene control region which is active in skeletal muscle (Sani, 1985, *Nature* 314:283-86); and the 30 gonadotropic releasing hormone gene control region which is active in the hypothalamus (Mason *et al.*, 1986, *Science* 234:1372-78).

An enhancer sequence may be inserted into the vector to increase the transcription of a DNA encoding a GPCR polypeptide of the present invention by higher eukaryotes. Enhancers are cis-acting elements of DNA, usually about 10-300

bp in length, that act on the promoter to increase transcription. Enhancers are relatively orientation and position independent. They have been found 5' and 3' to the transcription unit. Several enhancer sequences available from mammalian genes are known (*e.g.*, globin, elastase, albumin, alpha-feto-protein and insulin). Typically, however, an enhancer from a virus will be used. The SV40 enhancer, the cytomegalovirus early promoter enhancer, the polyoma enhancer, and adenovirus enhancers are exemplary enhancing elements for the activation of eukaryotic promoters. While an enhancer may be spliced into the vector at a position 5' or 3' to a GPCR nucleic acid molecule, it is typically located at a site 5' from the promoter.

Expression vectors of the invention may be constructed from a starting vector such as a commercially available vector. Such vectors may or may not contain all of the desired flanking sequences. Where one or more of the flanking sequences described herein are not already present in the vector, they may be individually obtained and ligated into the vector. Methods used for obtaining each of the flanking sequences are well known to one skilled in the art.

Preferred vectors for practicing this invention are those that are compatible with bacterial, insect, and mammalian host cells. Such vectors include, *inter alia*, pCRII, pCR3, and pcDNA3.1 (Invitrogen, San Diego, CA), pBSII (Stratagene, La Jolla, CA), pET15 (Novagen, Madison, WI), pGEX (Pharmacia Biotech, Piscataway, NJ), pEGFP-N2 (Clontech, Palo Alto, CA), pETL (BlueBacII, Invitrogen), pDSR-alpha (International Pub. No. WO 90/14363) and pFastBacDual (Gibco-BRL, Grand Island, NY).

Additional suitable vectors include, but are not limited to, cosmids, plasmids, or modified viruses, but it will be appreciated that the vector system must be compatible with the selected host cell. Such vectors include, but are not limited to plasmids such as Bluescript<sup>®</sup> plasmid derivatives (a high copy number ColE1-based phagemid; Stratagene Cloning Systems, La Jolla CA), PCR cloning plasmids designed for cloning Taq-amplified PCR products (*e.g.*, TOPO<sup>™</sup> TA Cloning<sup>®</sup> Kit and PCR2.1<sup>®</sup> plasmid derivatives; Invitrogen), and mammalian, yeast or virus vectors such as a baculovirus expression system (pBacPAK plasmid derivatives; Clontech).

After the vector has been constructed and a nucleic acid molecule encoding a GPCR polypeptide has been inserted into the proper site of the vector, the completed vector may be inserted into a suitable host cell for amplification and/or polypeptide

expression. The transformation of an expression vector for a GPCR polypeptide into a selected host cell may be accomplished by well known methods including methods such as transfection, infection, calcium chloride, electroporation, microinjection, lipofection, DEAE-dextran method, or other known techniques. The method selected  
5 will in part be a function of the type of host cell to be used. These methods and other suitable methods are well known to the skilled artisan, and are set forth, for example, in Sambrook *et al.*, *supra*.

Host cells may be prokaryotic host cells (such as *E. coli*) or eukaryotic host cells (such as a yeast, insect, or vertebrate cell). The host cell, when cultured under  
10 appropriate conditions, synthesizes a GPCR polypeptide that can subsequently be collected from the culture medium (if the host cell secretes it into the medium) or directly from the host cell producing it (if it is not secreted). The selection of an appropriate host cell will depend upon various factors, such as desired expression levels, polypeptide modifications that are desirable or necessary for activity (such as  
15 glycosylation or phosphorylation) and ease of folding into a biologically active molecule.

A number of suitable host cells are known in the art and many are available from the American Type Culture Collection (ATCC), Manassas, VA. Examples include, but are not limited to, mammalian cells, such as Chinese hamster ovary cells  
20 (CHO), CHO DHFR(-) cells (Urlaub *et al.*, 1980, *Proc. Natl. Acad. Sci. U.S.A.* 97:4216-20), human embryonic kidney (HEK) 293 or 293T cells, or 3T3 cells. The selection of suitable mammalian host cells and methods for transformation, culture, amplification, screening, product production, and purification are known in the art. Other suitable mammalian cell lines, are the monkey COS-1 and COS-7 cell lines,  
25 and the CV-1 cell line. Further exemplary mammalian host cells include primate cell lines and rodent cell lines, including transformed cell lines. Normal diploid cells, cell strains derived from *in vitro* culture of primary tissue, as well as primary explants, are also suitable. Candidate cells may be genotypically deficient in the selection gene, or may contain a dominantly acting selection gene. Other suitable mammalian cell lines  
30 include but are not limited to, mouse neuroblastoma N2A cells, HeLa, mouse L-929 cells, 3T3 lines derived from Swiss, Balb-c or NIH mice, BHK or HaK hamster cell lines. Each of these cell lines is known by and available to those skilled in the art of protein expression.

Similarly useful as host cells suitable for the present invention are bacterial cells. For example, the various strains of *E. coli* (e.g., HB101, DH5 $\alpha$ , DH10, and MC1061) are well-known as host cells in the field of biotechnology. Various strains of *B. subtilis*, *Pseudomonas spp.*, other *Bacillus spp.*, *Streptomyces spp.*, and the like  
5 may also be employed in this method.

Many strains of yeast cells known to those skilled in the art are also available as host cells for the expression of the polypeptides of the present invention. Preferred yeast cells include, for example, *Saccharomyces cerevisiae* and *Pichia pastoris*.

Additionally, where desired, insect cell systems may be utilized in the  
10 methods of the present invention. Such systems are described, for example, in Kitts *et al.*, 1993, *Biotechniques*, 14:810-17; Lucklow, 1993, *Curr. Opin. Biotechnol.* 4:564-72; and Lucklow *et al.*, 1993, *J. Virol.*, 67:4566-79. Preferred insect cells are Sf-9 and Hi5 (Invitrogen).

One may also use transgenic animals to express glycosylated GPCR  
15 polypeptides. For example, one may use a transgenic milk-producing animal (a cow or goat, for example) and obtain the present glycosylated polypeptide in the animal milk. One may also use plants to produce GPCR polypeptides, however, in general, the glycosylation occurring in plants is different from that produced in mammalian cells, and may result in a glycosylated product which is not suitable for human  
20 therapeutic use.

#### Polypeptide Production

Host cells comprising a GPCR polypeptide expression vector may be cultured using standard media well known to the skilled artisan. The media will usually  
25 contain all nutrients necessary for the growth and survival of the cells. Suitable media for culturing *E. coli* cells include, for example, Luria Broth (LB) and/or Terrific Broth (TB). Suitable media for culturing eukaryotic cells include Roswell Park Memorial Institute medium 1640 (RPMI 1640), Minimal Essential Medium (MEM) and/or Dulbecco's Modified Eagle Medium (DMEM), all of which may be  
30 supplemented with serum and/or growth factors as necessary for the particular cell line being cultured. A suitable medium for insect cultures is Grace's medium supplemented with yeastolate, lactalbumin hydrolysate, and/or fetal calf serum as necessary.

Typically, an antibiotic or other compound useful for selective growth of transfected or transformed cells is added as a supplement to the media. The compound to be used will be dictated by the selectable marker element present on the plasmid with which the host cell was transformed. For example, where the selectable  
5 marker element is kanamycin resistance, the compound added to the culture medium will be kanamycin. Other compounds for selective growth include ampicillin, tetracycline, and neomycin.

The amount of a GPCR polypeptide produced by a host cell can be evaluated using standard methods known in the art. Such methods include, without limitation,  
10 Western blot analysis, SDS-polyacrylamide gel electrophoresis, non-denaturing gel electrophoresis, High Performance Liquid Chromatography (HPLC) separation, immunoprecipitation, and/or activity assays such as DNA binding gel shift assays.

If a GPCR polypeptide has been designed to be secreted from the host cells, the majority of polypeptide may be found in the cell culture medium. If however, the  
15 GPCR polypeptide is not secreted from the host cells, it will be present in the cytoplasm and/or the nucleus (for eukaryotic host cells) or in the cytosol (for gram-negative bacteria host cells).

For a GPCR polypeptide situated in the host cell cytoplasm and/or nucleus (for eukaryotic host cells) or in the cytosol (for bacterial host cells), the intracellular  
20 material (including inclusion bodies for gram-negative bacteria) can be extracted from the host cell using any standard technique known to the skilled artisan. For example, the host cells can be lysed to release the contents of the periplasm/cytoplasm by French press, homogenization, and/or sonication followed by centrifugation.

If a GPCR polypeptide has formed inclusion bodies in the cytosol, the  
25 inclusion bodies can often bind to the inner and/or outer cellular membranes and thus will be found primarily in the pellet material after centrifugation. The pellet material can then be treated at pH extremes or with a chaotropic agent such as a detergent, guanidine, guanidine derivatives, urea, or urea derivatives in the presence of a reducing agent such as dithiothreitol at alkaline pH or tris carboxyethyl phosphine at  
30 acid pH to release, break apart, and solubilize the inclusion bodies. The solubilized GPCR polypeptide can then be analyzed using gel electrophoresis, immunoprecipitation, or the like. If it is desired to isolate the GPCR polypeptide, isolation may be accomplished using standard methods such as those described herein and in Marston *et al.*, 1990, *Meth. Enz.*, 182:264-75.

In some cases, a GPCR polypeptide may not be biologically active upon isolation. Various methods for "refolding" or converting the polypeptide to its tertiary structure and generating disulfide linkages can be used to restore biological activity. Such methods include exposing the solubilized polypeptide to a pH usually  
5 above 7 and in the presence of a particular concentration of a chaotrope. The selection of chaotrope is very similar to the choices used for inclusion body solubilization, but usually the chaotrope is used at a lower concentration and is not necessarily the same as chaotropes used for the solubilization. In most cases the refolding/oxidation solution will also contain a reducing agent or the reducing agent  
10 plus its oxidized form in a specific ratio to generate a particular redox potential allowing for disulfide shuffling to occur in the formation of the protein's cysteine bridges. Some of the commonly used redox couples include cysteine/cystamine, glutathione (GSH)/dithiobis GSH, cupric chloride, dithiothreitol (DTT)/dithiane DTT, and 2-2-mercaptoethanol (bME)/dithio-b(ME). In many instances, a cosolvent may be  
15 used or may be needed to increase the efficiency of the refolding, and the more common reagents used for this purpose include glycerol, polyethylene glycol of various molecular weights, arginine and the like.

If inclusion bodies are not formed to a significant degree upon expression of a GPCR polypeptide, then the polypeptide will be found primarily in the supernatant  
20 after centrifugation of the cell homogenate. The polypeptide may be further isolated from the supernatant using methods such as those described herein.

The purification of a GPCR polypeptide from solution can be accomplished using a variety of techniques. If the polypeptide has been synthesized such that it contains a tag such as Hexahistidine (GPCR polypeptide/hexaHis) or other small  
25 peptide such as FLAG (Eastman Kodak Co., New Haven, CT) or *myc* (Invitrogen) at either its carboxyl- or amino-terminus, it may be purified in a one-step process by passing the solution through an affinity column where the column matrix has a high affinity for the tag.

For example, polyhistidine binds with great affinity and specificity to nickel.  
30 Thus, an affinity column of nickel (such as the Qiagen<sup>®</sup> nickel columns) can be used for purification of GPCR polypeptide/polyHis. See, e.g., *Current Protocols in Molecular Biology* § 10.11.8 (Ausubel *et al.*, eds., Green Publishers Inc. and Wiley and Sons 1993).

Additionally, GPCR polypeptides may be purified through the use of a monoclonal antibody that is capable of specifically recognizing and binding to a GPCR polypeptide.

Other suitable procedures for purification include, without limitation, affinity  
5 chromatography, immunoaffinity chromatography, ion exchange chromatography, molecular sieve chromatography, HPLC, electrophoresis (including native gel electrophoresis) followed by gel elution, and preparative isoelectric focusing ("Isoprime" machine/technique, Hoefer Scientific, San Francisco, CA). In some cases, two or more purification techniques may be combined to achieve increased  
10 purity.

GPCR polypeptides may also be prepared by chemical synthesis methods (such as solid phase peptide synthesis) using techniques known in the art such as those set forth by Merrifield *et al.*, 1963, *J. Am. Chem. Soc.* 85:2149; Houghten *et al.*, 1985, *Proc Natl Acad. Sci. USA* 82:5132; and Stewart and Young, *Solid Phase*  
15 *Peptide Synthesis* (Pierce Chemical Co. 1984). Such polypeptides may be synthesized with or without a methionine on the amino-terminus. Chemically synthesized GPCR polypeptides may be oxidized using methods set forth in these references to form disulfide bridges. Chemically synthesized GPCR polypeptides are expected to have comparable biological activity to the corresponding GPCR  
20 polypeptides produced recombinantly or purified from natural sources, and thus may be used interchangeably with a recombinant or natural GPCR polypeptide.

Another means of obtaining GPCR polypeptide is via purification from biological samples such as source tissues and/or fluids in which the GPCR polypeptide is naturally found. Such purification can be conducted using methods for  
25 protein purification as described herein. The presence of the GPCR polypeptide during purification may be monitored, for example, using an antibody prepared against recombinantly produced GPCR polypeptide or peptide fragments thereof.

A number of additional methods for producing nucleic acids and polypeptides are known in the art, and the methods can be used to produce polypeptides having  
30 specificity for GPCR polypeptide. See, e.g., Roberts *et al.*, 1997, *Proc. Natl. Acad. Sci. U.S.A.* 94:12297-303, which describes the production of fusion proteins between an mRNA and its encoded peptide. See also, Roberts, 1999, *Curr. Opin. Chem. Biol.* 3:268-73. Additionally, U.S. Patent No. 5,824,469 describes methods for obtaining oligonucleotides capable of carrying out a specific biological function. The procedure



involves generating a heterogeneous pool of oligonucleotides, each having a 5' randomized sequence, a central preselected sequence, and a 3' randomized sequence. The resulting heterogeneous pool is introduced into a population of cells that do not exhibit the desired biological function. Subpopulations of the cells are then screened  
5 for those that exhibit a predetermined biological function. From that subpopulation, oligonucleotides capable of carrying out the desired biological function are isolated.

U.S. Patent Nos. 5,763,192; 5,814,476; 5,723,323; and 5,817,483 describe processes for producing peptides or polypeptides. This is done by producing stochastic genes or fragments thereof, and then introducing these genes into host cells  
10 which produce one or more proteins encoded by the stochastic genes. The host cells are then screened to identify those clones producing peptides or polypeptides having the desired activity.

Another method for producing peptides or polypeptides is described in International Pub. No. WO99/15650, filed by Athersys, Inc. Known as "Random  
15 Activation of Gene Expression for Gene Discovery" (RAGE-GD), the process involves the activation of endogenous gene expression or over-expression of a gene by *in situ* recombination methods. For example, expression of an endogenous gene is activated or increased by integrating a regulatory sequence into the target cell that is capable of activating expression of the gene by non-homologous or illegitimate  
20 recombination. The target DNA is first subjected to radiation, and a genetic promoter inserted. The promoter eventually locates a break at the front of a gene, initiating transcription of the gene. This results in expression of the desired peptide or polypeptide.

It will be appreciated that these methods can also be used to create  
25 comprehensive GPCR polypeptide expression libraries, which can subsequently be used for high throughput phenotypic screening in a variety of assays, such as biochemical assays, cellular assays, and whole organism assays (*e.g.*, plant, mouse, etc.).

### 30 Synthesis

It will be appreciated by those skilled in the art that the nucleic acid and polypeptide molecules described herein may be produced by recombinant and other means.

### Selective Binding Agents

The term “selective binding agent” refers to a molecule that has specificity for one or more GPCR polypeptides. Suitable selective binding agents include, but are not limited to, antibodies and derivatives thereof, polypeptides, and small molecules.

5 Suitable selective binding agents may be prepared using methods known in the art. An exemplary GPCR polypeptide selective binding agent of the present invention is capable of binding a certain portion of the GPCR polypeptide thereby inhibiting the binding of the polypeptide to a GPCR polypeptide receptor.

Selective binding agents such as antibodies and antibody fragments that bind  
10 GPCR polypeptides are within the scope of the present invention. The antibodies may be polyclonal including monospecific polyclonal; monoclonal (MAbs); recombinant; chimeric; humanized, such as complementarity-determining region (CDR)-grafted; human; single chain; and/or bispecific; as well as fragments; variants; or derivatives thereof. Antibody fragments include those portions of the antibody that bind to an  
15 epitope on the GPCR polypeptide. Examples of such fragments include Fab and F(ab') fragments generated by enzymatic cleavage of full-length antibodies. Other binding fragments include those generated by recombinant DNA techniques, such as the expression of recombinant plasmids containing nucleic acid sequences encoding antibody variable regions.

20 Polyclonal antibodies directed toward a GPCR polypeptide generally are produced in animals (e.g., rabbits or mice) by means of multiple subcutaneous or intraperitoneal injections of GPCR polypeptide and an adjuvant. It may be useful to conjugate a GPCR polypeptide to a carrier protein that is immunogenic in the species to be immunized, such as keyhole limpet hemocyanin, serum, albumin, bovine  
25 thyroglobulin, or soybean trypsin inhibitor. Also, aggregating agents such as alum are used to enhance the immune response. After immunization, the animals are bled and the serum is assayed for anti-GPCR antibody titer.

Monoclonal antibodies directed toward GPCR polypeptides are produced using any method that provides for the production of antibody molecules by  
30 continuous cell lines in culture. Examples of suitable methods for preparing monoclonal antibodies include the hybridoma methods of Kohler *et al.*, 1975, *Nature* 256:495-97 and the human B-cell hybridoma method (Kozbor, 1984, *J. Immunol.* 133:3001; Brodeur *et al.*, *Monoclonal Antibody Production Techniques and Applications* 51-63 (Marcel Dekker, Inc., 1987). Also provided by the invention are

hybridoma cell lines that produce monoclonal antibodies reactive with GPCR polypeptides.

Monoclonal antibodies of the invention may be modified for use as therapeutics. One embodiment is a "chimeric" antibody in which a portion of the heavy (H) and/or light (L) chain is identical with or homologous to a corresponding sequence in antibodies derived from a particular species or belonging to a particular antibody class or subclass, while the remainder of the chain(s) is/are identical with or homologous to a corresponding sequence in antibodies derived from another species or belonging to another antibody class or subclass. Also included are fragments of such antibodies, so long as they exhibit the desired biological activity. See U.S. Patent No. 4,816,567; Morrison *et al.*, 1985, *Proc. Natl. Acad. Sci.* 81:6851-55.

In another embodiment, a monoclonal antibody of the invention is a "humanized" antibody. Methods for humanizing non-human antibodies are well known in the art. See U.S. Patent Nos. 5,585,089 and 5,693,762. Generally, a humanized antibody has one or more amino acid residues introduced into it from a source that is non-human. Humanization can be performed, for example, using methods described in the art (Jones *et al.*, 1986, *Nature* 321:522-25; Riechmann *et al.*, 1998, *Nature* 332:323-27; Verhoeven *et al.*, 1988, *Science* 239:1534-36), by substituting at least a portion of a rodent complementarity-determining region for the corresponding regions of a human antibody.

Also encompassed by the invention are human antibodies that bind GPCR polypeptides. Using transgenic animals (*e.g.*, mice) that are capable of producing a repertoire of human antibodies in the absence of endogenous immunoglobulin production such antibodies are produced by immunization with a GPCR polypeptide antigen (*i.e.*, having at least 6 contiguous amino acids), optionally conjugated to a carrier. See, *e.g.*, Jakobovits *et al.*, 1993, *Proc. Natl. Acad. Sci.* 90:2551-55; Jakobovits *et al.*, 1993, *Nature* 362:255-58; Bruggermann *et al.*, 1993, *Year in Immuno.* 7:33. In one method, such transgenic animals are produced by incapacitating the endogenous loci encoding the heavy and light immunoglobulin chains therein, and inserting loci encoding human heavy and light chain proteins into the genome thereof. Partially modified animals (*i.e.*, those having less than the full complement of modifications) are then cross-bred to obtain an animal having all of the desired immune system modifications. When administered an immunogen, these transgenic animals produce antibodies with human (rather than, *e.g.*, murine) amino

acid sequences, including variable regions that are immunospecific for these antigens. See International App. Nos. PCT/US96/05928 and PCT/US93/06926. Additional methods are described in U.S. Patent No. 5,545,807, International App. Nos. PCT/US91/245 and PCT/GB89/01207, and in European Patent Nos. 546073B1 and  
5 546073A1. Human antibodies can also be produced by the expression of recombinant DNA in host cells or by expression in hybridoma cells as described herein.

In an alternative embodiment, human antibodies can also be produced from phage-display libraries (Hoogenboom *et al.*, 1991, *J. Mol. Biol.* 227:381; Marks *et al.*, 1991, *J. Mol. Biol.* 222:581). These processes mimic immune selection through  
10 the display of antibody repertoires on the surface of filamentous bacteriophage, and subsequent selection of phage by their binding to an antigen of choice. One such technique is described in International App. No. PCT/US98/17364, which describes the isolation of high affinity and functional agonistic antibodies for MPL- and msk-receptors using such an approach.

Chimeric, CDR grafted, and humanized antibodies are typically produced by recombinant methods. Nucleic acids encoding the antibodies are introduced into host cells and expressed using materials and procedures described herein. In a preferred embodiment, the antibodies are produced in mammalian host cells, such as CHO cells. Monoclonal (*e.g.*, human) antibodies may be produced by the expression of  
20 recombinant DNA in host cells or by expression in hybridoma cells as described herein.

The anti-GPCR antibodies of the invention may be employed in any known assay method, such as competitive binding assays, direct and indirect sandwich assays, and immunoprecipitation assays (Sola, *Monoclonal Antibodies: A Manual of*  
25 *Techniques* 147-158 (CRC Press, Inc., 1987)) for the detection and quantitation of GPCR polypeptides. The antibodies will bind GPCR polypeptides with an affinity that is appropriate for the assay method being employed.

For diagnostic applications, in certain embodiments, anti-GPCR antibodies may be labeled with a detectable moiety. The detectable moiety can be any one that  
30 is capable of producing, either directly or indirectly, a detectable signal. For example, the detectable moiety may be a radioisotope, such as  $^3\text{H}$ ,  $^{14}\text{C}$ ,  $^{32}\text{P}$ ,  $^{35}\text{S}$ ,  $^{125}\text{I}$ ,  $^{99}\text{Tc}$ ,  $^{111}\text{In}$ , or  $^{67}\text{Ga}$ ; a fluorescent or chemiluminescent compound, such as fluorescein isothiocyanate, rhodamine, or luciferin; or an enzyme, such as alkaline phosphatase,

$\beta$ -galactosidase, or horseradish peroxidase (Bayer, *et al.*, 1990, *Meth. Enz.* 184:138-63).

Competitive binding assays rely on the ability of a labeled standard (*e.g.*, a GPCR polypeptide, or an immunologically reactive portion thereof) to compete with the test sample analyte (an GPCR polypeptide) for binding with a limited amount of anti-GPCR antibody. The amount of a GPCR polypeptide in the test sample is inversely proportional to the amount of standard that becomes bound to the antibodies. To facilitate determining the amount of standard that becomes bound, the antibodies typically are insolubilized before or after the competition, so that the standard and analyte that are bound to the antibodies may conveniently be separated from the standard and analyte that remain unbound.

Sandwich assays typically involve the use of two antibodies, each capable of binding to a different immunogenic portion, or epitope, of the protein to be detected and/or quantitated. In a sandwich assay, the test sample analyte is typically bound by a first antibody that is immobilized on a solid support, and thereafter a second antibody binds to the analyte, thus forming an insoluble three-part complex. *See, e.g.*, U.S. Patent No. 4,376,110. The second antibody may itself be labeled with a detectable moiety (direct sandwich assays) or may be measured using an anti-immunoglobulin antibody that is labeled with a detectable moiety (indirect sandwich assays). For example, one type of sandwich assay is an enzyme-linked immunosorbent assay (ELISA), in which case the detectable moiety is an enzyme.

The selective binding agents, including anti-GPCR antibodies, are also useful for *in vivo* imaging. An antibody labeled with a detectable moiety may be administered to an animal, preferably into the bloodstream, and the presence and location of the labeled antibody in the host assayed. The antibody may be labeled with any moiety that is detectable in an animal, whether by nuclear magnetic resonance, radiology, or other detection means known in the art.

Selective binding agents of the invention, including antibodies, may be used as therapeutics. These therapeutic agents are generally agonists or antagonists, in that they either enhance or reduce, respectively, at least one of the biological activities of a GPCR polypeptide. In one embodiment, antagonist antibodies of the invention are antibodies or binding fragments thereof which are capable of specifically binding to a GPCR polypeptide and which are capable of inhibiting or eliminating the functional activity of a GPCR polypeptide *in vivo* or *in vitro*. In preferred embodiments, the

selective binding agent, *e.g.*, an antagonist antibody, will inhibit the functional activity of a GPCR polypeptide by at least about 50%, and preferably by at least about 80%. In another embodiment, the selective binding agent may be an anti-GPCR polypeptide antibody that is capable of interacting with a GPCR polypeptide binding partner (a ligand or receptor) thereby inhibiting or eliminating GPCR polypeptide activity *in vitro* or *in vivo*. Selective binding agents, including agonist and antagonist anti-GPCR polypeptide antibodies, are identified by screening assays that are well known in the art.

The invention also relates to a kit comprising GPCR selective binding agents (such as antibodies) and other reagents useful for detecting GPCR polypeptide levels in biological samples. Such reagents may include a detectable label, blocking serum, positive and negative control samples, and detection reagents.

#### Microarrays

It will be appreciated that DNA microarray technology can be utilized in accordance with the present invention. DNA microarrays are miniature, high-density arrays of nucleic acids positioned on a solid support, such as glass. Each cell or element within the array contains numerous copies of a single nucleic acid species that acts as a target for hybridization with a complementary nucleic acid sequence (*e.g.*, mRNA). In expression profiling using DNA microarray technology, mRNA is first extracted from a cell or tissue sample and then converted enzymatically to fluorescently labeled cDNA. This material is hybridized to the microarray and unbound cDNA is removed by washing. The expression of discrete genes represented on the array is then visualized by quantitating the amount of labeled cDNA that is specifically bound to each target nucleic acid molecule. In this way, the expression of thousands of genes can be quantitated in a high throughput, parallel manner from a single sample of biological material.

This high throughput expression profiling has a broad range of applications with respect to the GPCR molecules of the invention, including, but not limited to: the identification and validation of GPCR disease-related genes as targets for therapeutics; molecular toxicology of related GPCR molecules and inhibitors thereof; stratification of populations and generation of surrogate markers for clinical trials; and enhancing related GPCR polypeptide small molecule drug discovery by aiding in the identification of selective compounds in high throughput screens.

### Chemical Derivatives

Chemically modified derivatives of GPCR polypeptides may be prepared by one skilled in the art, given the disclosures described herein. GPCR polypeptide  
5 derivatives are modified in a manner that is different – either in the type or location of the molecules naturally attached to the polypeptide. Derivatives may include molecules formed by the deletion of one or more naturally-attached chemical groups. The polypeptide comprising the amino acid sequence of any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6, or other GPCR polypeptide, may be modified by the  
10 covalent attachment of one or more polymers. For example, the polymer selected is typically water-soluble so that the protein to which it is attached does not precipitate in an aqueous environment, such as a physiological environment. Included within the scope of suitable polymers is a mixture of polymers. Preferably, for therapeutic use of the end-product preparation, the polymer will be pharmaceutically acceptable.

15 The polymers each may be of any molecular weight and may be branched or unbranched. The polymers each typically have an average molecular weight of between about 2 kDa to about 100 kDa (the term “about” indicating that in preparations of a water-soluble polymer, some molecules will weigh more, some less, than the stated molecular weight). The average molecular weight of each polymer is  
20 preferably between about 5 kDa and about 50 kDa, more preferably between about 12 kDa and about 40 kDa and most preferably between about 20 kDa and about 35 kDa.

Suitable water-soluble polymers or mixtures thereof include, but are not limited to, N-linked or O-linked carbohydrates, sugars, phosphates, polyethylene glycol (PEG) (including the forms of PEG that have been used to derivatize proteins,  
25 including mono-(C<sub>1</sub>-C<sub>10</sub>), alkoxy-, or aryloxy-polyethylene glycol), monomethoxy-polyethylene glycol, dextran (such as low molecular weight dextran of, for example, about 6 kD), cellulose, or other carbohydrate based polymers, poly-(N-vinyl pyrrolidone) polyethylene glycol, propylene glycol homopolymers, polypropylene oxide/ethylene oxide co-polymers, polyoxyethylated polyols (e.g., glycerol), and  
30 polyvinyl alcohol. Also encompassed by the present invention are bifunctional crosslinking molecules that may be used to prepare covalently attached GPCR polypeptide multimers.

In general, chemical derivatization may be performed under any suitable condition used to react a protein with an activated polymer molecule. Methods for

preparing chemical derivatives of polypeptides will generally comprise the steps of: (a) reacting the polypeptide with the activated polymer molecule (such as a reactive ester or aldehyde derivative of the polymer molecule) under conditions whereby the polypeptide comprising the amino acid sequence of any of SEQ ID NO: 2, SEQ ID  
5 NO: 4, or SEQ ID NO: 6, or other GPCR polypeptide, becomes attached to one or more polymer molecules, and (b) obtaining the reaction products. The optimal reaction conditions will be determined based on known parameters and the desired result. For example, the larger the ratio of polymer molecules to protein, the greater the percentage of attached polymer molecule. In one embodiment, the GPCR  
10 polypeptide derivative may have a single polymer molecule moiety at the amino-terminus. *See, e.g.*, U.S. Patent No. 5,234,784.

The pegylation of a polypeptide may be specifically carried out using any of the pegylation reactions known in the art. Such reactions are described, for example, in the following references: Francis *et al.*, 1992, *Focus on Growth Factors* 3:4-10;  
15 European Patent Nos. 0154316 and 0401384; and U.S. Patent No. 4,179,337. For example, pegylation may be carried out via an acylation reaction or an alkylation reaction with a reactive polyethylene glycol molecule (or an analogous reactive water-soluble polymer) as described herein. For the acylation reactions, a selected polymer should have a single reactive ester group. For reductive alkylation, a selected  
20 polymer should have a single reactive aldehyde group. A reactive aldehyde is, for example, polyethylene glycol propionaldehyde, which is water stable, or mono C<sub>1</sub>-C<sub>10</sub> alkoxy or aryloxy derivatives thereof (*see* U.S. Patent No. 5,252,714).

In another embodiment, GPCR polypeptides may be chemically coupled to biotin. The biotin/GPCR polypeptide molecules are then allowed to bind to avidin,  
25 resulting in tetravalent avidin/biotin/GPCR polypeptide molecules. GPCR polypeptides may also be covalently coupled to dinitrophenol (DNP) or trinitrophenol (TNP) and the resulting conjugates precipitated with anti-DNP or anti-TNP-IgM to form decameric conjugates with a valency of 10.

Generally, conditions that may be alleviated or modulated by the  
30 administration of the present GPCR polypeptide derivatives include those described herein for GPCR polypeptides. However, the GPCR polypeptide derivatives disclosed herein may have additional activities, enhanced or reduced biological activity, or other characteristics, such as increased or decreased half-life, as compared to the non-derivatized molecules.



### Genetically Engineered Non-Human Animals

Additionally included within the scope of the present invention are non-human animals such as mice, rats, or other rodents; rabbits, goats, sheep, or other farm  
5 animals, in which the genes encoding native GPCR polypeptide have been disrupted (*i.e.*, “knocked out”) such that the level of expression of GPCR polypeptide is significantly decreased or completely abolished. Such animals may be prepared using techniques and methods such as those described in U.S. Patent No. 5,557,032.

The present invention further includes non-human animals such as mice, rats,  
10 or other rodents; rabbits, goats, sheep, or other farm animals, in which either the native form of a GPCR gene for that animal or a heterologous GPCR gene is over-expressed by the animal, thereby creating a “transgenic” animal. Such transgenic animals may be prepared using well known methods such as those described in U.S. Patent No 5,489,743 and International Pub. No. WO 94/28122.

15 The present invention further includes non-human animals in which the promoter for one or more of the GPCR polypeptides of the present invention is either activated or inactivated (*e.g.*, by using homologous recombination methods) to alter the level of expression of one or more of the native GPCR polypeptides.

These non-human animals may be used for drug candidate screening. In such  
20 screening, the impact of a drug candidate on the animal may be measured. For example, drug candidates may decrease or increase the expression of the GPCR gene. In certain embodiments, the amount of GPCR polypeptide that is produced may be measured after the exposure of the animal to the drug candidate. Additionally, in certain embodiments, one may detect the actual impact of the drug candidate on the  
25 animal. For example, over-expression of a particular gene may result in, or be associated with, a disease or pathological condition. In such cases, one may test a drug candidate’s ability to decrease expression of the gene or its ability to prevent or inhibit a pathological condition. In other examples, the production of a particular metabolic product such as a fragment of a polypeptide, may result in, or be associated  
30 with, a disease or pathological condition. In such cases, one may test a drug candidate’s ability to decrease the production of such a metabolic product or its ability to prevent or inhibit a pathological condition.

### Assaying for Other Modulators of GPCR Polypeptide Activity

In some situations, it may be desirable to identify molecules that are modulators, *i.e.*, agonists or antagonists, of the activity of GPCR polypeptide. Natural or synthetic molecules that modulate GPCR polypeptide may be identified using one or more screening assays, such as those described herein. Such molecules may be administered either in an *ex vivo* manner or in an *in vivo* manner by injection, or by oral delivery, implantation device, or the like.

“Test molecule” refers to a molecule that is under evaluation for the ability to modulate (*i.e.*, increase or decrease) the activity of a GPCR polypeptide. Most commonly, a test molecule will interact directly with a GPCR polypeptide. However, it is also contemplated that a test molecule may also modulate GPCR polypeptide activity indirectly, such as by affecting GPCR gene expression, or by binding to a GPCR polypeptide binding partner (*e.g.*, receptor or ligand). In one embodiment, a test molecule will bind to a GPCR polypeptide with an affinity constant of at least about  $10^{-6}$  M, preferably about  $10^{-8}$  M, more preferably about  $10^{-9}$  M, and even more preferably about  $10^{-10}$  M.

Methods for identifying compounds that interact with GPCR polypeptides are encompassed by the present invention. In certain embodiments, a GPCR polypeptide is incubated with a test molecule under conditions that permit the interaction of the test molecule with a GPCR polypeptide, and the extent of the interaction is measured. The test molecule can be screened in a substantially purified form or in a crude mixture.

In certain embodiments, a GPCR polypeptide agonist or antagonist may be a protein, peptide, carbohydrate, lipid, or small molecular weight molecule that interacts with GPCR polypeptide to regulate its activity. Molecules which regulate GPCR polypeptide expression include nucleic acids which are complementary to nucleic acids encoding a GPCR polypeptide, or are complementary to nucleic acids sequences which direct or control the expression of GPCR polypeptide, and which act as anti-sense regulators of expression.

Once a test molecule has been identified as interacting with a GPCR polypeptide, the molecule may be further evaluated for its ability to increase or decrease GPCR polypeptide activity. The measurement of the interaction of a test molecule with GPCR polypeptide may be carried out in several formats, including cell-based binding assays, membrane binding assays, solution-phase assays, and immunoassays. In general, a test molecule is incubated with a GPCR polypeptide for

a specified period of time, and GPCR polypeptide activity is determined by one or more assays for measuring biological activity.

The interaction of test molecules with GPCR polypeptides may also be assayed directly using polyclonal or monoclonal antibodies in an immunoassay.

- 5 Alternatively, modified forms of GPCR polypeptides containing epitope tags as described herein may be used in solution and immunoassays.

In the event that GPCR polypeptides display biological activity through an interaction with a binding partner (e.g., a receptor or a ligand), a variety of *in vitro* assays may be used to measure the binding of a GPCR polypeptide to the  
10 corresponding binding partner (such as a selective binding agent, receptor, or ligand). These assays may be used to screen test molecules for their ability to increase or decrease the rate and/or the extent of binding of a GPCR polypeptide to its binding partner. In one assay, a GPCR polypeptide is immobilized in the wells of a microtiter plate. Radiolabeled GPCR polypeptide binding partner (for example, iodinated  
15 GPCR polypeptide binding partner) and a test molecule can then be added either one at a time (in either order) or simultaneously to the wells. After incubation, the wells can be washed and counted for radioactivity, using a scintillation counter, to determine the extent to which the binding partner bound to the GPCR polypeptide. Typically, a molecule will be tested over a range of concentrations, and a series of  
20 control wells lacking one or more elements of the test assays can be used for accuracy in the evaluation of the results. An alternative to this method involves reversing the "positions" of the proteins, i.e., immobilizing GPCR polypeptide binding partner to the microtiter plate wells, incubating with the test molecule and radiolabeled GPCR polypeptide, and determining the extent of GPCR polypeptide binding. See, e.g.,  
25 *Current Protocols in Molecular Biology*, chap. 18 (Ausubel *et al.*, eds., Green Publishers Inc. and Wiley and Sons 1995).

As an alternative to radiolabeling, a GPCR polypeptide or its binding partner may be conjugated to biotin, and the presence of biotinylated protein can then be detected using streptavidin linked to an enzyme, such as horse radish peroxidase  
30 (HRP) or alkaline phosphatase (AP), which can be detected colorometrically, or by fluorescent tagging of streptavidin. An antibody directed to a GPCR polypeptide or to a GPCR polypeptide binding partner, and which is conjugated to biotin, may also be used for purposes of detection following incubation of the complex with enzyme-linked streptavidin linked to AP or HRP.

A GPCR polypeptide or a GPCR polypeptide binding partner can also be immobilized by attachment to agarose beads, acrylic beads, or other types of such inert solid phase substrates. The substrate-protein complex can be placed in a solution containing the complementary protein and the test compound. After incubation, the  
5 beads can be precipitated by centrifugation, and the amount of binding between a GPCR polypeptide and its binding partner can be assessed using the methods described herein. Alternatively, the substrate-protein complex can be immobilized in a column with the test molecule and complementary protein passing through the column. The formation of a complex between a GPCR polypeptide and its binding  
10 partner can then be assessed using any of the techniques described herein (*e.g.*, radiolabelling or antibody binding).

Another *in vitro* assay that is useful for identifying a test molecule that increases or decreases the formation of a complex between a GPCR polypeptide binding protein and a GPCR polypeptide binding partner is a surface plasmon  
15 resonance detector system such as the BIAcore assay system (Pharmacia, Piscataway, NJ). The BIAcore system is utilized as specified by the manufacturer. This assay essentially involves the covalent binding of either GPCR polypeptide or a GPCR polypeptide binding partner to a dextran-coated sensor chip that is located in a detector. The test compound and the other complementary protein can then be  
20 injected, either simultaneously or sequentially, into the chamber containing the sensor chip. The amount of complementary protein that binds can be assessed based on the change in molecular mass that is physically associated with the dextran-coated side of the sensor chip, with the change in molecular mass being measured by the detector system.

25 In some cases, it may be desirable to evaluate two or more test compounds together for their ability to increase or decrease the formation of a complex between a GPCR polypeptide and a GPCR polypeptide binding partner. In these cases, the assays set forth herein can be readily modified by adding such additional test compound(s) either simultaneously with, or subsequent to, the first test compound.  
30 The remainder of the steps in the assay are as set forth herein.

*In vitro* assays such as those described herein may be used advantageously to screen large numbers of compounds for an effect on the formation of a complex between a GPCR polypeptide and GPCR polypeptide binding partner. The assays

may be automated to screen compounds generated in phage display, synthetic peptide, and chemical synthesis libraries.

Compounds which increase or decrease the formation of a complex between a GPCR polypeptide and a GPCR polypeptide binding partner may also be screened in  
5 cell culture using cells and cell lines expressing either GPCR polypeptide or GPCR polypeptide binding partner. Cells and cell lines may be obtained from any mammal, but preferably will be from human or other primate, canine, or rodent sources. The binding of a GPCR polypeptide to cells expressing GPCR polypeptide binding partner at the surface is evaluated in the presence or absence of test molecules, and the extent  
10 of binding may be determined by, for example, flow cytometry using a biotinylated antibody to a GPCR polypeptide binding partner. Cell culture assays can be used advantageously to further evaluate compounds that score positive in protein binding assays described herein.

Cell cultures can also be used to screen the impact of a drug candidate. For  
15 example, drug candidates may decrease or increase the expression of the GPCR gene. In certain embodiments, the amount of GPCR polypeptide or a GPCR polypeptide fragment that is produced may be measured after exposure of the cell culture to the drug candidate. In certain embodiments, one may detect the actual impact of the drug candidate on the cell culture. For example, the over-expression of a particular gene  
20 may have a particular impact on the cell culture. In such cases, one may test a drug candidate's ability to increase or decrease the expression of the gene or its ability to prevent or inhibit a particular impact on the cell culture. In other examples, the production of a particular metabolic product such as a fragment of a polypeptide, may result in, or be associated with, a disease or pathological condition. In such cases, one  
25 may test a drug candidate's ability to decrease the production of such a metabolic product in a cell culture.

#### Internalizing Proteins

The *tat* protein sequence (from HIV) can be used to internalize proteins into a  
30 cell. See, e.g., Falwell *et al.*, 1994, *Proc. Natl. Acad. Sci. U.S.A.* 91:664-68. For example, an 11 amino acid sequence (Y-G-R-K-K-R-R-Q-R-R-R; SEQ ID NO: 7) of the HIV *tat* protein (termed the "protein transduction domain," or TAT PDT) has been described as mediating delivery across the cytoplasmic membrane and the nuclear membrane of a cell. See Schwarze *et al.*, 1999, *Science* 285:1569-72; and Nagahara

*et al.*, 1998, *Nat. Med.* 4:1449-52. In these procedures, FITC-constructs (FITC-labeled G-G-G-G-Y-G-R-K-K-R-R-Q-R-R-R; SEQ ID NO: 8), which penetrate tissues following intraperitoneal administration, are prepared, and the binding of such constructs to cells is detected by fluorescence-activated cell sorting (FACS) analysis.

5 Cells treated with a *tat*- $\beta$ -gal fusion protein will demonstrate  $\beta$ -gal activity. Following injection, expression of such a construct can be detected in a number of tissues, including liver, kidney, lung, heart, and brain tissue. It is believed that such constructs undergo some degree of unfolding in order to enter the cell, and as such, may require a refolding following entry into the cell.

10 It will thus be appreciated that the *tat* protein sequence may be used to internalize a desired polypeptide into a cell. For example, using the *tat* protein sequence, a GPCR antagonist (such as an anti-GPCR selective binding agent, small molecule, soluble receptor, or antisense oligonucleotide) can be administered intracellularly to inhibit the activity of a GPCR molecule. As used herein, the term  
15 "GPCR molecule" refers to both GPCR nucleic acid molecules and GPCR polypeptides as defined herein. Where desired, the GPCR protein itself may also be internally administered to a cell using these procedures. *See also*, Straus, 1999, *Science* 285:1466-67.

#### 20 Cell Source Identification Using GPCR Polypeptide

In accordance with certain embodiments of the invention, it may be useful to be able to determine the source of a certain cell type associated with a GPCR polypeptide. For example, it may be useful to determine the origin of a disease or pathological condition as an aid in selecting an appropriate therapy. In certain  
25 embodiments, nucleic acids encoding a GPCR polypeptide can be used as a probe to identify cells described herein by screening the nucleic acids of the cells with such a probe. In other embodiments, one may use anti-GPCR polypeptide antibodies to test for the presence of GPCR polypeptide in cells, and thus, determine if such cells are of the types described herein.

30

#### GPCR Polypeptide Compositions and Administration

Therapeutic compositions are within the scope of the present invention. Such GPCR polypeptide pharmaceutical compositions may comprise a therapeutically effective amount of a GPCR polypeptide or a GPCR nucleic acid molecule in

admixture with a pharmaceutically or physiologically acceptable formulation agent selected for suitability with the mode of administration. Pharmaceutical compositions may comprise a therapeutically effective amount of one or more GPCR polypeptide selective binding agents in admixture with a pharmaceutically or physiologically acceptable formulation agent selected for suitability with the mode of administration.

Acceptable formulation materials preferably are nontoxic to recipients at the dosages and concentrations employed.

The pharmaceutical composition may contain formulation materials for modifying, maintaining, or preserving, for example, the pH, osmolarity, viscosity, clarity, color, isotonicity, odor, sterility, stability, rate of dissolution or release, adsorption, or penetration of the composition. Suitable formulation materials include, but are not limited to, amino acids (such as glycine, glutamine, asparagine, arginine, or lysine), antimicrobials, antioxidants (such as ascorbic acid, sodium sulfite, or sodium hydrogen-sulfite), buffers (such as borate, bicarbonate, Tris-HCl, citrates, phosphates, or other organic acids), bulking agents (such as mannitol or glycine), chelating agents (such as ethylenediamine tetraacetic acid (EDTA)), complexing agents (such as caffeine, polyvinylpyrrolidone, beta-cyclodextrin, or hydroxypropyl-beta-cyclodextrin), fillers, monosaccharides, disaccharides, and other carbohydrates (such as glucose, mannose, or dextrans), proteins (such as serum albumin, gelatin, or immunoglobulins), coloring, flavoring and diluting agents, emulsifying agents, hydrophilic polymers (such as polyvinylpyrrolidone), low molecular weight polypeptides, salt-forming counterions (such as sodium), preservatives (such as benzalkonium chloride, benzoic acid, salicylic acid, thimerosal, phenethyl alcohol, methylparaben, propylparaben, chlorhexidine, sorbic acid, or hydrogen peroxide), solvents (such as glycerin, propylene glycol, or polyethylene glycol), sugar alcohols (such as mannitol or sorbitol), suspending agents, surfactants or wetting agents (such as pluronics; PEG; sorbitan esters; polysorbates such as polysorbate 20 or polysorbate 80; triton; tromethamine; lecithin; cholesterol or tyloxapal), stability enhancing agents (such as sucrose or sorbitol), tonicity enhancing agents (such as alkali metal halides – preferably sodium or potassium chloride – or mannitol sorbitol), delivery vehicles, diluents, excipients and/or pharmaceutical adjuvants. *See Remington's Pharmaceutical Sciences* (18th Ed., A.R. Gennaro, ed., Mack Publishing Company 1990.

The optimal pharmaceutical composition will be determined by a skilled

artisan depending upon, for example, the intended route of administration, delivery format, and desired dosage. *See, e.g., Remington's Pharmaceutical Sciences, supra.* Such compositions may influence the physical state, stability, rate of *in vivo* release, and rate of *in vivo* clearance of the GPCR molecule.

5           The primary vehicle or carrier in a pharmaceutical composition may be either aqueous or non-aqueous in nature. For example, a suitable vehicle or carrier for injection may be water, physiological saline solution, or artificial cerebrospinal fluid, possibly supplemented with other materials common in compositions for parenteral administration. Neutral buffered saline or saline mixed with serum albumin are  
10 further exemplary vehicles. Other exemplary pharmaceutical compositions comprise Tris buffer of about pH 7.0-8.5, or acetate buffer of about pH 4.0-5.5, which may further include sorbitol or a suitable substitute. In one embodiment of the present invention, GPCR polypeptide compositions may be prepared for storage by mixing the selected composition having the desired degree of purity with optional  
15 formulation agents (*Remington's Pharmaceutical Sciences, supra*) in the form of a lyophilized cake or an aqueous solution. Further, the GPCR polypeptide product may be formulated as a lyophilizate using appropriate excipients such as sucrose.

          The GPCR polypeptide pharmaceutical compositions can be selected for parenteral delivery. Alternatively, the compositions may be selected for inhalation or  
20 for delivery through the digestive tract, such as orally. The preparation of such pharmaceutically acceptable compositions is within the skill of the art.

          The formulation components are present in concentrations that are acceptable to the site of administration. For example, buffers are used to maintain the composition at physiological pH or at a slightly lower pH, typically within a pH range  
25 of from about 5 to about 8.

          When parenteral administration is contemplated, the therapeutic compositions for use in this invention may be in the form of a pyrogen-free, parenterally acceptable, aqueous solution comprising the desired GPCR molecule in a pharmaceutically acceptable vehicle. A particularly suitable vehicle for parenteral injection is sterile  
30 distilled water in which a GPCR molecule is formulated as a sterile, isotonic solution, properly preserved. Yet another preparation can involve the formulation of the desired molecule with an agent, such as injectable microspheres, bio-erodible particles, polymeric compounds (such as polylactic acid or polyglycolic acid), beads, or liposomes, that provides for the controlled or sustained release of the product



which may then be delivered via a depot injection. Hyaluronic acid may also be used, and this may have the effect of promoting sustained duration in the circulation. Other suitable means for the introduction of the desired molecule include implantable drug delivery devices.

5           In one embodiment, a pharmaceutical composition may be formulated for inhalation. For example, GPCR polypeptide may be formulated as a dry powder for inhalation. GPCR polypeptide or nucleic acid molecule inhalation solutions may also be formulated with a propellant for aerosol delivery. In yet another embodiment, solutions may be nebulized. Pulmonary administration is further described in  
10   International Pub. No. WO 94/20069, which describes the pulmonary delivery of chemically modified proteins.

          It is also contemplated that certain formulations may be administered orally. In one embodiment of the present invention, GPCR polypeptides that are administered in this fashion can be formulated with or without those carriers customarily used in  
15   the compounding of solid dosage forms such as tablets and capsules. For example, a capsule may be designed to release the active portion of the formulation at the point in the gastrointestinal tract when bioavailability is maximized and pre-systemic degradation is minimized. Additional agents can be included to facilitate absorption of the GPCR polypeptide. Diluents, flavorings, low melting point waxes, vegetable  
20   oils, lubricants, suspending agents, tablet disintegrating agents, and binders may also be employed.

          Another pharmaceutical composition may involve an effective quantity of GPCR polypeptides in a mixture with non-toxic excipients that are suitable for the manufacture of tablets. By dissolving the tablets in sterile water, or another  
25   appropriate vehicle, solutions can be prepared in unit-dose form. Suitable excipients include, but are not limited to, inert diluents, such as calcium carbonate, sodium carbonate or bicarbonate, lactose, or calcium phosphate; or binding agents, such as starch, gelatin, or acacia; or lubricating agents such as magnesium stearate, stearic acid, or talc.

30           Additional GPCR polypeptide pharmaceutical compositions will be evident to those skilled in the art, including formulations involving GPCR polypeptides in sustained- or controlled-delivery formulations. Techniques for formulating a variety of other sustained- or controlled-delivery means, such as liposome carriers, bio-erodible microparticles or porous beads and depot injections, are also known to those

skilled in the art. *See, e.g.*, International App. No. PCT/US93/00829, which describes the controlled release of porous polymeric microparticles for the delivery of pharmaceutical compositions.

Additional examples of sustained-release preparations include semipermeable  
5 polymer matrices in the form of shaped articles, *e.g.* films, or microcapsules. Sustained release matrices may include polyesters, hydrogels, polylactides (U.S. Patent No. 3,773,919 and European Patent No. 058481), copolymers of L-glutamic acid and gamma ethyl-L-glutamate (Sidman *et al.*, 1983, *Biopolymers* 22:547-56), poly(2-hydroxyethyl-methacrylate) (Langer *et al.*, 1981, *J. Biomed. Mater. Res.*  
10 15:167-277 and Langer, 1982, *Chem. Tech.* 12:98-105), ethylene vinyl acetate (Langer *et al.*, *supra*) or poly-D(-)-3-hydroxybutyric acid (European Patent No. 133988). Sustained-release compositions may also include liposomes, which can be prepared by any of several methods known in the art. *See, e.g.*, Eppstein *et al.*, 1985, *Proc. Natl. Acad. Sci. USA* 82:3688-92; and European Patent Nos. 036676, 088046,  
15 and 143949.

The GPCR pharmaceutical composition to be used for *in vivo* administration typically must be sterile. This may be accomplished by filtration through sterile filtration membranes. Where the composition is lyophilized, sterilization using this method may be conducted either prior to, or following, lyophilization and  
20 reconstitution. The composition for parenteral administration may be stored in lyophilized form or in a solution. In addition, parenteral compositions generally are placed into a container having a sterile access port, for example, an intravenous solution bag or vial having a stopper pierceable by a hypodermic injection needle.

Once the pharmaceutical composition has been formulated, it may be stored in  
25 sterile vials as a solution, suspension, gel, emulsion, solid, or as a dehydrated or lyophilized powder. Such formulations may be stored either in a ready-to-use form or in a form (*e.g.*, lyophilized) requiring reconstitution prior to administration.

In a specific embodiment, the present invention is directed to kits for producing a single-dose administration unit. The kits may each contain both a first  
30 container having a dried protein and a second container having an aqueous formulation. Also included within the scope of this invention are kits containing single and multi-chambered pre-filled syringes (*e.g.*, liquid syringes and lyosyringes).

The effective amount of a GPCR pharmaceutical composition to be employed therapeutically will depend, for example, upon the therapeutic context and objectives.

One skilled in the art will appreciate that the appropriate dosage levels for treatment will thus vary depending, in part, upon the molecule delivered, the indication for which the GPCR molecule is being used, the route of administration, and the size (body weight, body surface, or organ size) and condition (the age and general health) of the patient. Accordingly, the clinician may titer the dosage and modify the route of administration to obtain the optimal therapeutic effect. A typical dosage may range from about 0.1  $\mu\text{g/kg}$  to up to about 100 mg/kg or more, depending on the factors mentioned above. In other embodiments, the dosage may range from 0.1  $\mu\text{g/kg}$  up to about 100 mg/kg; or 1  $\mu\text{g/kg}$  up to about 100 mg/kg; or 5  $\mu\text{g/kg}$  up to about 100 mg/kg.

The frequency of dosing will depend upon the pharmacokinetic parameters of the GPCR molecule in the formulation being used. Typically, a clinician will administer the composition until a dosage is reached that achieves the desired effect. The composition may therefore be administered as a single dose, as two or more doses (which may or may not contain the same amount of the desired molecule) over time, or as a continuous infusion via an implantation device or catheter. Further refinement of the appropriate dosage is routinely made by those of ordinary skill in the art and is within the ambit of tasks routinely performed by them. Appropriate dosages may be ascertained through use of appropriate dose-response data.

The route of administration of the pharmaceutical composition is in accord with known methods, *e.g.*, orally; through injection by intravenous, intraperitoneal, intracerebral (intraparenchymal), intracerebroventricular, intramuscular, intraocular, intraarterial, intraportal, or intralesional routes; by sustained release systems; or by implantation devices. Where desired, the compositions may be administered by bolus injection or continuously by infusion, or by implantation device.

Alternatively or additionally, the composition may be administered locally via implantation of a membrane, sponge, or other appropriate material onto which the desired molecule has been absorbed or encapsulated. Where an implantation device is used, the device may be implanted into any suitable tissue or organ, and delivery of the desired molecule may be via diffusion, timed-release bolus, or continuous administration.

In some cases, it may be desirable to use GPCR polypeptide pharmaceutical compositions in an *ex vivo* manner. In such instances, cells, tissues, or organs that have been removed from the patient are exposed to GPCR polypeptide

pharmaceutical compositions after which the cells, tissues, or organs are subsequently implanted back into the patient.

In other cases, a GPCR polypeptide can be delivered by implanting certain cells that have been genetically engineered, using methods such as those described  
5 herein, to express and secrete the GPCR polypeptide. Such cells may be animal or human cells, and may be autologous, heterologous, or xenogeneic. Optionally, the cells may be immortalized. In order to decrease the chance of an immunological response, the cells may be encapsulated to avoid infiltration of surrounding tissues. The encapsulation materials are typically biocompatible, semi-permeable polymeric  
10 enclosures or membranes that allow the release of the protein product(s) but prevent the destruction of the cells by the patient's immune system or by other detrimental factors from the surrounding tissues.

As discussed herein, it may be desirable to treat isolated cell populations (such as stem cells, lymphocytes, red blood cells, chondrocytes, neurons, and the like) with  
15 one or more GPCR polypeptides. This can be accomplished by exposing the isolated cells to the polypeptide directly, where it is in a form that is permeable to the cell membrane.

Additional embodiments of the present invention relate to cells and methods (e.g., homologous recombination and/or other recombinant production methods) for  
20 both the *in vitro* production of therapeutic polypeptides and for the production and delivery of therapeutic polypeptides by gene therapy or cell therapy. Homologous and other recombination methods may be used to modify a cell that contains a normally transcriptionally-silent GPCR gene, or an under-expressed gene, and thereby produce a cell which expresses therapeutically efficacious amounts of GPCR  
25 polypeptides.

Homologous recombination is a technique originally developed for targeting genes to induce or correct mutations in transcriptionally active genes. Kucherlapati, 1989, *Prog. in Nucl. Acid Res. & Mol. Biol.* 36:301. The basic technique was developed as a method for introducing specific mutations into specific regions of the  
30 mammalian genome (Thomas *et al.*, 1986, *Cell* 44:419-28; Thomas and Capecchi, 1987, *Cell* 51:503-12; Doetschman *et al.*, 1988, *Proc. Natl. Acad. Sci. U.S.A.* 85:8583-87) or to correct specific mutations within defective genes (Doetschman *et al.*, 1987, *Nature* 330:576-78). Exemplary homologous recombination techniques are described

in U.S. Patent No. 5,272,071; European Patent Nos. 9193051 and 505500; International App. No. PCT/US90/07642, and International Pub No. WO 91/09955).

Through homologous recombination, the DNA sequence to be inserted into the genome can be directed to a specific region of the gene of interest by attaching it to  
5 targeting DNA. The targeting DNA is a nucleotide sequence that is complementary (homologous) to a region of the genomic DNA. Small pieces of targeting DNA that are complementary to a specific region of the genome are put in contact with the parental strand during the DNA replication process. It is a general property of DNA that has been inserted into a cell to hybridize, and therefore, recombine with other  
10 pieces of endogenous DNA through shared homologous regions. If this complementary strand is attached to an oligonucleotide that contains a mutation or a different sequence or an additional nucleotide, it too is incorporated into the newly synthesized strand as a result of the recombination. As a result of the proofreading function, it is possible for the new sequence of DNA to serve as the template. Thus,  
15 the transferred DNA is incorporated into the genome.

Attached to these pieces of targeting DNA are regions of DNA that may interact with or control the expression of a GPCR polypeptide, *e.g.*, flanking sequences. For example, a promoter/enhancer element, a suppressor, or an exogenous transcription modulatory element is inserted in the genome of the intended host cell in  
20 proximity and orientation sufficient to influence the transcription of DNA encoding the desired GPCR polypeptide. The control element controls a portion of the DNA present in the host cell genome. Thus, the expression of the desired GPCR polypeptide may be achieved not by transfection of DNA that encodes the GPCR gene itself, but rather by the use of targeting DNA (containing regions of homology with  
25 the endogenous gene of interest) coupled with DNA regulatory segments that provide the endogenous gene sequence with recognizable signals for transcription of a GPCR gene.

In an exemplary method, the expression of a desired targeted gene in a cell (*i.e.*, a desired endogenous cellular gene) is altered via homologous recombination  
30 into the cellular genome at a preselected site, by the introduction of DNA that includes at least a regulatory sequence, an exon, and a splice donor site. These components are introduced into the chromosomal (genomic) DNA in such a manner that this, in effect, results in the production of a new transcription unit (in which the regulatory sequence, the exon, and the splice donor site present in the DNA construct

are operatively linked to the endogenous gene). As a result of the introduction of these components into the chromosomal DNA, the expression of the desired endogenous gene is altered.

Altered gene expression, as described herein, encompasses activating (or  
5 causing to be expressed) a gene which is normally silent (unexpressed) in the cell as obtained, as well as increasing the expression of a gene which is not expressed at physiologically significant levels in the cell as obtained. The embodiments further encompass changing the pattern of regulation or induction such that it is different from the pattern of regulation or induction that occurs in the cell as obtained, and  
10 reducing (including eliminating) the expression of a gene which is expressed in the cell as obtained.

One method by which homologous recombination can be used to increase, or cause, GPCR polypeptide production from a cell's endogenous GPCR gene involves first using homologous recombination to place a recombination sequence from a site-specific recombination system (*e.g.*, Cre/loxP, FLP/FRT) (Sauer, 1994, *Curr. Opin. Biotechnol.*, 5:521-27; Sauer, 1993, *Methods Enzymol.*, 225:890-900) upstream of  
15 (*i.e.*, 5' to) the cell's endogenous genomic GPCR polypeptide coding region. A plasmid containing a recombination site homologous to the site that was placed just upstream of the genomic GPCR polypeptide coding region is introduced into the modified cell line along with the appropriate recombinase enzyme. This recombinase causes the plasmid to integrate, via the plasmid's recombination site, into the recombination site located just upstream of the genomic GPCR polypeptide coding region in the cell line (Baubonis and Sauer, 1993, *Nucleic Acids Res.* 21:2025-29; O'Gorman *et al.*, 1991, *Science* 251:1351-55). Any flanking sequences known to  
20 increase transcription (*e.g.*, enhancer/promoter, intron, translational enhancer), if properly positioned in this plasmid, would integrate in such a manner as to create a new or modified transcriptional unit resulting in *de novo* or increased GPCR polypeptide production from the cell's endogenous GPCR gene.

A further method to use the cell line in which the site specific recombination  
30 sequence had been placed just upstream of the cell's endogenous genomic GPCR polypeptide coding region is to use homologous recombination to introduce a second recombination site elsewhere in the cell line's genome. The appropriate recombinase enzyme is then introduced into the two-recombination-site cell line, causing a recombination event (deletion, inversion, and translocation) (Sauer, 1994, *Curr. Opin.*

*Biotechnol.*, 5:521-27; Sauer, 1993, *Methods Enzymol.*, 225:890-900) that would create a new or modified transcriptional unit resulting in *de novo* or increased GPCR polypeptide production from the cell's endogenous GPCR gene.

An additional approach for increasing, or causing, the expression of GPCR polypeptide from a cell's endogenous GPCR gene involves increasing, or causing, the expression of a gene or genes (*e.g.*, transcription factors) and/or decreasing the expression of a gene or genes (*e.g.*, transcriptional repressors) in a manner which results in *de novo* or increased GPCR polypeptide production from the cell's endogenous GPCR gene. This method includes the introduction of a non-naturally occurring polypeptide (*e.g.*, a polypeptide comprising a site specific DNA binding domain fused to a transcriptional factor domain) into the cell such that *de novo* or increased GPCR polypeptide production from the cell's endogenous GPCR gene results.

The present invention further relates to DNA constructs useful in the method of altering expression of a target gene. In certain embodiments, the exemplary DNA constructs comprise: (a) one or more targeting sequences, (b) a regulatory sequence, (c) an exon, and (d) an unpaired splice-donor site. The targeting sequence in the DNA construct directs the integration of elements (a) - (d) into a target gene in a cell such that the elements (b) - (d) are operatively linked to sequences of the endogenous target gene. In another embodiment, the DNA constructs comprise: (a) one or more targeting sequences, (b) a regulatory sequence, (c) an exon, (d) a splice-donor site, (e) an intron, and (f) a splice-acceptor site, wherein the targeting sequence directs the integration of elements (a) - (f) such that the elements of (b) - (f) are operatively linked to the endogenous gene. The targeting sequence is homologous to the preselected site in the cellular chromosomal DNA with which homologous recombination is to occur. In the construct, the exon is generally 3' of the regulatory sequence and the splice-donor site is 3' of the exon.

If the sequence of a particular gene is known, such as the nucleic acid sequence of GPCR polypeptide presented herein, a piece of DNA that is complementary to a selected region of the gene can be synthesized or otherwise obtained, such as by appropriate restriction of the native DNA at specific recognition sites bounding the region of interest. This piece serves as a targeting sequence upon insertion into the cell and will hybridize to its homologous region within the genome. If this hybridization occurs during DNA replication, this piece of DNA, and any

additional sequence attached thereto, will act as an Okazaki fragment and will be incorporated into the newly synthesized daughter strand of DNA. The present invention, therefore, includes nucleotides encoding a GPCR polypeptide, which nucleotides may be used as targeting sequences.

5 GPCR polypeptide cell therapy, *e.g.*, the implantation of cells producing GPCR polypeptides, is also contemplated. This embodiment involves implanting cells capable of synthesizing and secreting a biologically active form of GPCR polypeptide. Such GPCR polypeptide-producing cells can be cells that are natural producers of GPCR polypeptides or may be recombinant cells whose ability to  
10 produce GPCR polypeptides has been augmented by transformation with a gene encoding the desired GPCR polypeptide or with a gene augmenting the expression of GPCR polypeptide. Such a modification may be accomplished by means of a vector suitable for delivering the gene as well as promoting its expression and secretion. In order to minimize a potential immunological reaction in patients being administered a  
15 GPCR polypeptide, as may occur with the administration of a polypeptide of a foreign species, it is preferred that the natural cells producing GPCR polypeptide be of human origin and produce human GPCR polypeptide. Likewise, it is preferred that the recombinant cells producing GPCR polypeptide be transformed with an expression vector containing a gene encoding a human GPCR polypeptide.

20 Implanted cells may be encapsulated to avoid the infiltration of surrounding tissue. Human or non-human animal cells may be implanted in patients in biocompatible, semipermeable polymeric enclosures or membranes that allow the release of GPCR polypeptide, but that prevent the destruction of the cells by the patient's immune system or by other detrimental factors from the surrounding tissue.  
25 Alternatively, the patient's own cells, transformed to produce GPCR polypeptides *ex vivo*, may be implanted directly into the patient without such encapsulation.

Techniques for the encapsulation of living cells are known in the art, and the preparation of the encapsulated cells and their implantation in patients may be routinely accomplished. For example, Baetge *et al.* (International Pub. No. WO  
30 95/05452 and International App. No. PCT/US94/09299) describe membrane capsules containing genetically engineered cells for the effective delivery of biologically active molecules. The capsules are biocompatible and are easily retrievable. The capsules encapsulate cells transfected with recombinant DNA molecules comprising DNA sequences coding for biologically active molecules operatively linked to promoters



that are not subject to down-regulation *in vivo* upon implantation into a mammalian host. The devices provide for the delivery of the molecules from living cells to specific sites within a recipient. In addition, *see* U.S. Patent Nos. 4,892,538; 5,011,472; and 5,106,627. A system for encapsulating living cells is described in  
5 International Pub. No. WO 91/10425 (Aebischer *et al.*). *See also*, International Pub. No. WO 91/10470 (Aebischer *et al.*); Winn *et al.*, 1991, *Exper. Neurol.* 113:322-29; Aebischer *et al.*, 1991, *Exper. Neurol.* 111:269-75; and Tresco *et al.*, 1992, *ASAIO* 38:17-23.

*In vivo* and *in vitro* gene therapy delivery of GPCR polypeptides is also  
10 envisioned. One example of a gene therapy technique is to use the GPCR gene (either genomic DNA, cDNA, and/or synthetic DNA) encoding a GPCR polypeptide that may be operably linked to a constitutive or inducible promoter to form a "gene therapy DNA construct." The promoter may be homologous or heterologous to the endogenous GPCR gene, provided that it is active in the cell or tissue type into which  
15 the construct will be inserted. Other components of the gene therapy DNA construct may optionally include DNA molecules designed for site-specific integration (*e.g.*, endogenous sequences useful for homologous recombination), tissue-specific promoters, enhancers or silencers, DNA molecules capable of providing a selective advantage over the parent cell, DNA molecules useful as labels to identify  
20 transformed cells, negative selection systems, cell specific binding agents (as, for example, for cell targeting), cell-specific internalization factors, transcription factors enhancing expression from a vector, and factors enabling vector production.

A gene therapy DNA construct can then be introduced into cells (either *ex vivo* or *in vivo*) using viral or non-viral vectors. One means for introducing the gene  
25 therapy DNA construct is by means of viral vectors as described herein. Certain vectors, such as retroviral vectors, will deliver the DNA construct to the chromosomal DNA of the cells, and the gene can integrate into the chromosomal DNA. Other vectors will function as episomes, and the gene therapy DNA construct will remain in the cytoplasm.

30 In yet other embodiments, regulatory elements can be included for the controlled expression of the GPCR gene in the target cell. Such elements are turned on in response to an appropriate effector. In this way, a therapeutic polypeptide can be expressed when desired. One conventional control means involves the use of small molecule dimerizers or rapalogs to dimerize chimeric proteins which contain a small

molecule-binding domain and a domain capable of initiating a biological process, such as a DNA-binding protein or transcriptional activation protein (*see* International Pub. Nos. WO 96/41865, WO 97/31898, and WO 97/31899). The dimerization of the proteins can be used to initiate transcription of the transgene.

5           An alternative regulation technology uses a method of storing proteins expressed from the gene of interest inside the cell as an aggregate or cluster. The gene of interest is expressed as a fusion protein that includes a conditional aggregation domain that results in the retention of the aggregated protein in the endoplasmic reticulum. The stored proteins are stable and inactive inside the cell.  
10       The proteins can be released, however, by administering a drug (*e.g.*, small molecule ligand) that removes the conditional aggregation domain and thereby specifically breaks apart the aggregates or clusters so that the proteins may be secreted from the cell. *See* Aridor *et al.*, 2000, *Science* 287:816-17 and Rivera *et al.*, 2000, *Science* 287:826-30.

15           Other suitable control means or gene switches include, but are not limited to, the systems described herein. Mifepristone (RU486) is used as a progesterone antagonist. The binding of a modified progesterone receptor ligand-binding domain to the progesterone antagonist activates transcription by forming a dimer of two transcription factors that then pass into the nucleus to bind DNA. The ligand-binding  
20       domain is modified to eliminate the ability of the receptor to bind to the natural ligand. The modified steroid hormone receptor system is further described in U.S. Patent No. 5,364,791 and International Pub. Nos. WO 96/40911 and WO 97/10337.

          Yet another control system uses ecdysone (a fruit fly steroid hormone), which binds to and activates an ecdysone receptor (cytoplasmic receptor). The receptor then  
25       translocates to the nucleus to bind a specific DNA response element (promoter from ecdysone-responsive gene). The ecdysone receptor includes a transactivation domain, DNA-binding domain, and ligand-binding domain to initiate transcription. The ecdysone system is further described in U.S. Patent No. 5,514,578 and International Pub. Nos. WO 97/38117, WO 96/37609, and WO 93/03162.

30           Another control means uses a positive tetracycline-controllable transactivator. This system involves a mutated tet repressor protein DNA-binding domain (mutated tet R-4 amino acid changes which resulted in a reverse tetracycline-regulated transactivator protein, *i.e.*, it binds to a tet operator in the presence of tetracycline)

linked to a polypeptide which activates transcription. Such systems are described in U.S. Patent Nos. 5,464,758, 5,650,298, and 5,654,168.

Additional expression control systems and nucleic acid constructs are described in U.S. Patent Nos. 5,741,679 and 5,834,186, to Innovir Laboratories Inc.

5        *In vivo* gene therapy may be accomplished by introducing the gene encoding GPCR polypeptide into cells via local injection of a GPCR nucleic acid molecule or by other appropriate viral or non-viral delivery vectors. Hefti 1994, *Neurobiology* 25:1418-35. For example, a nucleic acid molecule encoding a GPCR polypeptide may be contained in an adeno-associated virus (AAV) vector for delivery to the  
10        targeted cells (*see, e.g.*, Johnson, International Pub. No. WO 95/34670; International App. No. PCT/US95/07178). The recombinant AAV genome typically contains AAV inverted terminal repeats flanking a DNA sequence encoding a GPCR polypeptide operably linked to functional promoter and polyadenylation sequences.

Alternative suitable viral vectors include, but are not limited to, retrovirus,  
15        adenovirus, herpes simplex virus, lentivirus, hepatitis virus, parvovirus, papovavirus, poxvirus, alphavirus, coronavirus, rhabdovirus, paramyxovirus, and papilloma virus vectors. U.S. Patent No. 5,672,344 describes an *in vivo* viral-mediated gene transfer system involving a recombinant neurotrophic HSV-1 vector. U.S. Patent No. 5,399,346 provides examples of a process for providing a patient with a therapeutic  
20        protein by the delivery of human cells that have been treated *in vitro* to insert a DNA segment encoding a therapeutic protein. Additional methods and materials for the practice of gene therapy techniques are described in U.S. Patent Nos. 5,631,236 (involving adenoviral vectors), 5,672,510 (involving retroviral vectors), 5,635,399 (involving retroviral vectors expressing cytokines).

25        Nonviral delivery methods include, but are not limited to, liposome-mediated transfer, naked DNA delivery (direct injection), receptor-mediated transfer (ligand-DNA complex), electroporation, calcium phosphate precipitation, and microparticle bombardment (*e.g.*, gene gun). Gene therapy materials and methods may also include inducible promoters, tissue-specific enhancer-promoters, DNA sequences designed for  
30        site-specific integration, DNA sequences capable of providing a selective advantage over the parent cell, labels to identify transformed cells, negative selection systems and expression control systems (safety measures), cell-specific binding agents (for cell targeting), cell-specific internalization factors, and transcription factors to enhance expression by a vector as well as methods of vector manufacture. Such additional

methods and materials for the practice of gene therapy techniques are described in U.S. Patent Nos. 4,970,154 (involving electroporation techniques), 5,679,559 (describing a lipoprotein-containing system for gene delivery), 5,676,954 (involving liposome carriers), 5,593,875 (describing methods for calcium phosphate  
5 transfection), and 4,945,050 (describing a process wherein biologically active particles are propelled at cells at a speed whereby the particles penetrate the surface of the cells and become incorporated into the interior of the cells), and International Pub. No. WO 96/40958 (involving nuclear ligands).

It is also contemplated that GPCR gene therapy or cell therapy can further  
10 include the delivery of one or more additional polypeptide(s) in the same or a different cell(s). Such cells may be separately introduced into the patient, or the cells may be contained in a single implantable device, such as the encapsulating membrane described above, or the cells may be separately modified by means of viral vectors.

A means to increase endogenous GPCR polypeptide expression in a cell via  
15 gene therapy is to insert one or more enhancer elements into the GPCR polypeptide promoter, where the enhancer elements can serve to increase transcriptional activity of the GPCR gene. The enhancer elements used will be selected based on the tissue in which one desires to activate the gene – enhancer elements known to confer promoter activation in that tissue will be selected. For example, if a gene encoding a  
20 GPCR polypeptide is to be “turned on” in T-cells, the *lck* promoter enhancer element may be used. Here, the functional portion of the transcriptional element to be added may be inserted into a fragment of DNA containing the GPCR polypeptide promoter (and optionally, inserted into a vector and/or 5’ and/or 3’ flanking sequences) using standard cloning techniques. This construct, known as a “homologous recombination  
25 construct,” can then be introduced into the desired cells either *ex vivo* or *in vivo*.

Gene therapy also can be used to decrease GPCR polypeptide expression by modifying the nucleotide sequence of the endogenous promoter. Such modification is typically accomplished via homologous recombination methods. For example, a DNA molecule containing all or a portion of the promoter of the GPCR gene selected  
30 for inactivation can be engineered to remove and/or replace pieces of the promoter that regulate transcription. For example, the TATA box and/or the binding site of a transcriptional activator of the promoter may be deleted using standard molecular biology techniques; such deletion can inhibit promoter activity thereby repressing the transcription of the corresponding GPCR gene. The deletion of the TATA box or the

transcription activator binding site in the promoter may be accomplished by generating a DNA construct comprising all or the relevant portion of the GPCR polypeptide promoter (from the same or a related species as the GPCR gene to be regulated) in which one or more of the TATA box and/or transcriptional activator binding site nucleotides are mutated via substitution, deletion and/or insertion of one or more nucleotides. As a result, the TATA box and/or activator binding site has decreased activity or is rendered completely inactive. This construct, which also will typically contain at least about 500 bases of DNA that correspond to the native (endogenous) 5' and 3' DNA sequences adjacent to the promoter segment that has been modified, may be introduced into the appropriate cells (either *ex vivo* or *in vivo*) either directly or via a viral vector as described herein. Typically, the integration of the construct into the genomic DNA of the cells will be via homologous recombination, where the 5' and 3' DNA sequences in the promoter construct can serve to help integrate the modified promoter region via hybridization to the endogenous chromosomal DNA.

#### Therapeutic Uses

GPCR nucleic acid molecules, polypeptides, and agonists and antagonists thereof can be used to treat, diagnose, ameliorate, or prevent a number of diseases, disorders, or conditions, including those recited herein.

GPCR polypeptide agonists and antagonists include those molecules which regulate GPCR polypeptide activity and either increase or decrease at least one activity of the mature form of the GPCR polypeptide. Agonists or antagonists may be co-factors, such as a protein, peptide, carbohydrate, lipid, or small molecular weight molecule, which interact with GPCR polypeptide and thereby regulate its activity. Potential polypeptide agonists or antagonists include antibodies that react with either soluble or membrane-bound forms of GPCR polypeptides that comprise part or all of the extracellular domains of the said proteins. Molecules that regulate GPCR polypeptide expression typically include nucleic acids encoding GPCR polypeptide that can act as anti-sense regulators of expression.

Since GPCR polypeptide expression has been detected in the fat tissue, GPCR nucleic acid molecules, polypeptides, agonists and antagonists thereof (including, but not limited to, anti-GPCR selective binding agents) may be useful for the treatment or diagnosis of diseases involving fat metabolism. Examples of such diseases include,

but are not limited to, obesity, diabetes, and aberrant lipid metabolism (for example, dislipidemia). The molecules of the present invention may also have application to the treatment of weight loss associated with cancer (*i.e.*, cachexia) and other conditions of abnormal weight loss, such as those relating to AIDS, anorexia nervosa.

5 Other diseases associated with fat metabolism are encompassed within the scope of the invention.

GPCR polypeptides may also function biologically as mediators of inflammation. Molecules which regulate GPCR expression or activation may therefore be useful as agents for the treatment of related diseases or abnormal  
10 conditions, including but not necessarily limited to inflammatory diseases mediated by leukotrienes, such as pulmonary disorders (for example, asthma, chronic bronchitis, and related obstructive airway diseases), allergies and allergic reactions (for example, allergic rhinitis, contact dermatitis, and allergic conjunctivitis), angina, cerebral spasm, glomerular nephritis, hepatitis, endotoxemia, uveitis, and allograft  
15 rejection.

Since GPCR polypeptide expression has been detected in the testes, GPCR nucleic acid molecules, polypeptides, and agonists and antagonists thereof may be useful in diagnosing or treating diseases and conditions affecting the testes. Examples of such diseases and conditions include, but are not limited to, male infertility and  
20 testicular carcinoma. Other diseases and conditions associated with the development and function of the testes are encompassed within the scope of this invention.

More broadly, the GPCR molecules of the present invention may provide methods of treating other abnormal conditions related to an excess or insufficient amount of GPCR (signaling) activity. In general, if the activity of the GPCR is in  
25 excess, several approaches are available. One approach will comprise administering to a subject an inhibitor compound (antagonist) along with a pharmaceutically acceptable carrier in an amount effective to inhibit activation by blocking binding of the ligand to the GPCR, or by inhibiting a second signal and thereby alleviating the abnormal condition. In another approach, expression of the gene encoding the  
30 endogenous GPCR can be inhibited using expression blocking techniques, such as by antisense sequences either internally generated or separately administered.

Agonists or antagonists of GPCR polypeptide function may be used (simultaneously or sequentially) in combination with one or more cytokines, growth

factors, antibiotics, anti-inflammatories, and/or chemotherapeutic agents as is appropriate for the condition being treated.

Other diseases or disorders caused by or mediated by undesirable levels of GPCR polypeptides are encompassed within the scope of the invention. Undesirable  
5 levels include excessive levels of GPCR polypeptides and sub-normal levels of GPCR polypeptides.

#### Uses of GPCR Nucleic Acids and Polypeptides

Nucleic acid molecules of the invention (including those that do not  
10 themselves encode biologically active polypeptides) may be used to map the locations of the GPCR gene and related genes on chromosomes. Mapping may be done by techniques known in the art, such as PCR amplification and *in situ* hybridization.

GPCR nucleic acid molecules (including those that do not themselves encode biologically active polypeptides), may be useful as hybridization probes in diagnostic  
15 assays to test, either qualitatively or quantitatively, for the presence of a GPCR nucleic acid molecule in mammalian tissue or bodily fluid samples.

Other methods may also be employed where it is desirable to inhibit the activity of one or more GPCR polypeptides. Such inhibition may be effected by nucleic acid molecules that are complementary to and hybridize to expression control  
20 sequences (triple helix formation) or to GPCR mRNA. For example, antisense DNA or RNA molecules, which have a sequence that is complementary to at least a portion of a GPCR gene can be introduced into the cell. Anti-sense probes may be designed by available techniques using the sequence of the GPCR gene disclosed herein. Typically, each such antisense molecule will be complementary to the start site (5'  
25 end) of each selected GPCR gene. When the antisense molecule then hybridizes to the corresponding GPCR mRNA, translation of this mRNA is prevented or reduced. Anti-sense inhibitors provide information relating to the decrease or absence of a GPCR polypeptide in a cell or organism.

Alternatively, gene therapy may be employed to create a dominant-negative  
30 inhibitor of one or more GPCR polypeptides. In this situation, the DNA encoding a mutant polypeptide of each selected GPCR polypeptide can be prepared and introduced into the cells of a patient using either viral or non-viral methods as described herein. Each such mutant is typically designed to compete with endogenous polypeptide in its biological role.

In addition, a GPCR polypeptide, whether biologically active or not, may be used as an immunogen, that is, the polypeptide contains at least one epitope to which antibodies may be raised. Selective binding agents that bind to a GPCR polypeptide (as described herein) may be used for *in vivo* and *in vitro* diagnostic purposes, including, but not limited to, use in labeled form to detect the presence of GPCR polypeptide in a body fluid or cell sample. The antibodies may also be used to prevent, treat, or diagnose a number of diseases and disorders, including those recited herein. The antibodies may bind to a GPCR polypeptide so as to diminish or block at least one activity characteristic of a GPCR polypeptide, or may bind to a polypeptide to increase at least one activity characteristic of a GPCR polypeptide (including by increasing the pharmacokinetics of the GPCR polypeptide).

GPCR polypeptides can be used to clone GPCR ligands using an "expression cloning" strategy. Radiolabeled (<sup>125</sup>Iodine) GPCR polypeptide or "affinity/activity-tagged" GPCR polypeptide (such as an Fc fusion or an alkaline phosphatase fusion) can be used in binding assays to identify a cell type, cell line, or tissue that expresses a GPCR ligand. RNA isolated from such cells or tissues can then be converted to cDNA, cloned into a mammalian expression vector, and transfected into mammalian cells (*e.g.*, COS or 293) to create an expression library. Radiolabeled or tagged GPCR polypeptide can then be used as an affinity reagent to identify and isolate the subset of cells in this library expressing a GPCR ligand. DNA is then isolated from these cells and transfected into mammalian cells to create a secondary expression library in which the fraction of cells expressing the GPCR ligand would be many-fold higher than in the original library. This enrichment process can be repeated iteratively until a single recombinant clone containing the GPCR ligand is isolated. Isolation of GPCR ligands is useful for identifying or developing novel agonists and antagonists of the GPCR signaling pathway. Such agonists and antagonists include GPCR ligands, anti-GPCR ligand antibodies, small molecules or antisense oligonucleotides.

The human GPCR nucleic acids of the present invention are also useful tools for isolating the corresponding chromosomal GPCR polypeptide genes. The human GPCR genomic DNA can be used to identify heritable tissue-degenerating diseases.

The following examples are intended for illustration purposes only, and should not be construed as limiting the scope of the invention in any way.

#### Example 1: Cloning of the Rat GPCR Polypeptide Gene



Generally, materials and methods as described in Sambrook *et al. supra* were used to clone and analyze genes encoding rat GPCR polypeptides.

To isolate cDNA sequences encoding rat GPCR polypeptide, a homology-based search of a proprietary expressed sequence tag (EST) database was performed.

5 A partial clone from a white adipose tissue cDNA library was found to encode a potential G-protein coupled receptor. To generate the full-length cDNA sequence for GPCR polypeptide, RACE was performed using a rat adipocyte Marathon cDNA RACE kit (Clontech) and the conditions suggested by the manufacturer. Two successive PCR reactions were carried out. In the first reaction (PCR1), the AP1  
10 primer from the Marathon cDNA RACE kit was used with the primer 5'-A-A-G-A-G-G-A-C-C-A-G-G-C-G-G-C-A-G-G-G-A-A-T-A-T-3' (SEQ ID NO: 9). In the second reaction (PCR2), the AP2 primer from the kit was used with the primer 5'-T-A-T-C-C-C-C-A-A-A-T-C-C-A-A-T-G-C-C-T-A-C-G-3' (SEQ ID NO: 10). The resulting RACE DNA was ligated into the PCRII vector (Invitrogen) and transformed  
15 into *E. coli*. Positive clones were sequenced, and a consensus sequence was constructed. A full-length clone was obtained using a high fidelity PCR and primers derived from the sequence determined by RACE. Amplification reactions contained rat adipocyte cDNA (prepared with the Marathon cDNA RACE kit), the primers 5'-C-G-G-G-C-A-G-G-T-G-G-G-T-G-A-T-G-A-G-G-T-T-A-G-3' (SEQ ID NO: 11) and 5'-  
20 G-C-T-G-C-T-G-G-G-C-C-A-T-T-T-G-T-C-T-T-C-A-T-3' (SEQ ID NO: 12), 10  $\mu$ L reaction buffer (100 mM Tricine, pH 5.7, 25% glycerol, and 425 mM KOAc), 4  $\mu$ L of dNTPs (1mM each), 1  $\mu$ L of rtTh (Perkin Elmer, 2 U/ $\mu$ L), and 1  $\mu$ L of Vent polymerase (New England Biolabs, 0.01 U/ $\mu$ L), in a total volume of 49  $\mu$ L. Reactions were performed at 94°C for 10 seconds, 62°C for 1 minute, and 68°C for 5 minutes for  
25 35 cycles. When the temperature of the thermal cycler reached 65°C, 1  $\mu$ L of 10 mM MgOAc was added. The PCR product obtained was analyzed by agarose gel electrophoresis, and the appropriate fragment was then excised from the gel and purified using a GENECLAN kit (Bio101; Vista, CA). The purified DNA was cloned using the PCRscript kit (Stratagene; La Jolla, CA) and a clone containing the  
30 amplified DNA was recovered in *E. coli*. The DNA sequence of the insert was then determined.

A full-length rat GPCR cDNA clone was also isolated from an oligo-dT primed cDNA library constructed from rat white adipose tissue. Total RNA was isolated from female Wistar rat white adipose tissue using TRIzol Reagent (Life

Technologies; Rockville, MD). Messenger RNA was purified from the total RNA using Dynal Dynabeads. The cDNA was synthesized and cloned into the vector, pSPORT1, using 2.7 µg of mRNA and the Superscript Plasmid System for cDNA synthesis and plasmid cloning (Life Technologies). The cDNA was introduced into DH10B *E. coli* (Life Technologies) by electroporation. CsCl gradient-purified plasmid DNA was harvested from a 100 ml Terrific Broth culture containing over 1 million unamplified cDNA clones. Using the Genetrappor system (Life Technologies), 5 µg of CsCl gradient purified plasmid cDNA was screened with the primer 5'-T-G-C-T-G-T-C-T-C-A-T-C-G-A-G-G--G-G-G-A-A-3' (SEQ ID NO: 13). One-fifth of the rescued plasmid was electroporated in DH10B *E. coli* cells and then plated onto LB agar plates containing 100 µg/ml of ampicillin. Colony hybridization was performed using Hybond N+ nylon membranes (Amersham; Piscataway, NJ) and an end-labeled primer (5'-G-A-A-T-A-G-G-G-C-C-G-G-A-A-G-C-A-T-T-G-T-3'; SEQ ID NO: 14). Hybridized colonies were rescreened by PCR using the primers 5'-C-C-T-C-C-T-C-A-T-C-C-G-A-G-C-C-T-G-T-C-T-G-G-3' (SEQ ID NO: 15) and 5'-C-C-T-T-T-G-T-G-T-C-A-G-C-C-A-C-C-T-A-G-G-A-T-G-C-3' (SEQ ID NO: 16). One positive isolated colony was grown overnight in LB containing 100 µg/ml of ampicillin. Plasmid DNA was isolated using a Plasmid Maxi Kit (Qiagen), and sequenced. Sequence analysis of the full-length cDNA for murine GPCR polypeptide indicated that the murine GPCR gene comprises a 1053 bp open reading frame encoding a protein of 351 amino acids (Figure 3A-3C).

#### Example 2: Cloning of the Murine GPCR Polypeptide Gene

Generally, materials and methods as described in Sambrook *et al. supra* were used to clone and analyze genes encoding murine GPCR polypeptides.

To isolate cDNA sequences encoding murine GPCR polypeptide, the rat GPCR cDNA sequence was used to search a proprietary genomics database. A cDNA clone from a mouse bone marrow cDNA library was identified that matched the 5' end of the rat clone. The clone was recovered and its complete sequence was determined. Sequence analysis of the full-length cDNA for rat GPCR polypeptide indicated that the rat GPCR gene comprises a 1053 bp open reading frame encoding a protein of 351 amino acids (Figures 2A-2D).

#### Example 3: Cloning of the Human GPCR Polypeptide Gene

Generally, materials and methods as described in Sambrook *et al. supra* were used to clone and analyze genes encoding human GPCR polypeptides.

To isolate cDNA sequences encoding human GPCR polypeptide, a <sup>32</sup>P-dCTP labeled 700 bp rat GPCR probe (prepared by PCR from the rat GPCR cDNA sequence using the primers 5'-C-C-T-C-C-T-C-A-T-C-C-G-A-G-C-C-T-G-T-C-T-G-G-3'; SEQ ID NO: 15 and 5'-C-C-T-T-T-G-T-G-T-C-A-G-C-C-A-C-C-T-A-G-G-A-T-G-C-3'; SEQ ID NO: 16) was used to screen a human white adipose tissue cDNA phage library (Clontech) by plaque hybridization. Recombinant phages were plated onto the *E. coli* strain XL1-blue at approximately 5 x 10<sup>4</sup> transformants per 150 mm LB plate. Positively charged nylon filters were lifted and prehybridized in 5x SSC, 5x Denhardt's solution, 0.5% SDS, and 200 µg/ml denatured salmon sperm DNA for 3 hours at 45°C. The filters were then hybridized overnight at a temperature of 45°C in the same solution supplemented with 5 ng/ml of labeled probe. The filters were first washed twice in 2x SSC and 0.1% SDS for 20 minutes at room temperature, and then the filters were washed twice in 0.5x SSC and 0.1% SDS at 45°C for 30 minutes. The filters were then exposed to x-ray film with intensifying screens at a temperature of -80°C for 5 days. Positive hybridizing plaques were determined by aligning to duplicate filters. Thirty-six positive plaques were picked and re-plated for a secondary screen using the same methods employed in the primary screening. The secondary screening produced 16 positive plaques.

Recombinant phage identified in the second screen was transduced into the *E. coli* strain BM25.8, which expresses cre-recombinase. The pTriplEx phagemid was excised as a result of cre-recombinase-mediated site-specific recombination at the loxP sites. The infected BM25.8 cells were plated on LB/ampicillin plates and incubated overnight. Positively charged 82 mm filters (NEN; Boston, MA) were lifted as described, and a tertiary screen was performed using the 700 bp labeled rat GPCR probe described herein. Pre-hybridization and hybridization were performed as described herein. Overnight exposure showed that all but two plates contained nearly 100% positive colonies. Two colonies per positive plate were amplified in LB media containing 100 mg/ml ampicillin. The plasmid DNA was prepared and both strands of cDNA insert were sequenced. DNA encoding human GPCR was found to be 2.5 kb in length. Sequence analysis of the full-length cDNA for human GPCR polypeptide indicated that the human GPCR gene comprises a 1038 bp open reading

frame encoding a protein of 346 amino acids (Figures 1A-1B).

Figure 4 illustrates an amino acid sequence alignment of human GPCR polypeptide (hu\_GPCR; SEQ ID NO: 2), murine GPCR polypeptide (mu\_GPCR; SEQ ID NO: 4) and rat GPCR polypeptide (ra\_GPCR; SEQ ID NO: 6). The human GPCR polypeptide shares a 78% identity with both the murine and rat GPCR polypeptides. The murine and rat GPCR polypeptides share a 90% identity. The structure of the human, murine, and rat GPCR polypeptides parallels that of other members of the G-protein coupled receptor family in that each polypeptide possesses seven transmembrane domains.

10

#### Example 4: GPCR mRNA Expression

Quantitative PCR using the PRISM Taqman system was performed to assess GPCR mRNA expression in multiple human cDNA samples (human poly A<sup>+</sup> was obtained from Clontech Laboratories; cDNA was prepared using the SuperScript Amplification System; Gibco BRL). PRISM Taqman reactions were performed using two primers and a fluorogenic probe that were derived from the human GPCR sequence. The data generated was then normalized to the human housekeeper gene cyclophilin and reported as a ratio of GPCR copy number to human cyclophilin copy number. Reactions for the quantitation of human GPCR mRNA expression contained the primers 5'-T-T-C-A-C-G-T-T-G-G-C-C-A-T-G-A-A-C-A-3' (SEQ ID NO: 17) and 5'-A-A-A-T-A-C-C-T-G-T-C-C-G-C-A-G-C-C-3' (SEQ ID NO: 18), and the fluorogenic probe 5'-(6-FAM)-C-C-G-T-A-A-G-G-A-A-C-A-C-G-A-T-G-C-T-C-C-C-G-G-(TAMRA)-3' (SEQ ID NO: 19; wherein "6-FAM" is the 5' reporter dye 6-carboxyfluorescein and "TAMRA" is the 3' quencher 6-carboxytetramethylrhodamine). Reactions for the quantitation of human cyclophilin mRNA expression contained the primers 5'-G-T-C-G-A-C-G-G-C-G-A-G-C-C-C-3' (SEQ ID NO: 20) and 5'-T-C-T-T-T-G-G-G-A-C-C-T-T-G-T-C-T-G-C-3' (SEQ ID NO: 21), and the fluorogenic probe 5'-(6-FAM)-T-G-G-G-C-C-G-C-G-T-C-T-C-C-T-T-T-G-A-G-C-T-(TAMRA)-3' (SEQ ID NO: 22). The highest levels of GPCR mRNA expression were detected in testes and white adipose tissue (Figure 5).

30

The expression of GPCR mRNA is examined by Northern blot analysis. Multiple human tissue northern blots (Clontech) are probed with a suitable restriction fragment isolated from a human GPCR polypeptide cDNA clone. The probe is

labeled with  $^{32}\text{P}$ -dCTP using standard techniques.

Northern blots are prehybridized for 2 hours at 42°C in hybridization solution (5X SSC, 50% deionized formamide, 5X Denhardt's solution, 0.5% SDS, and 100 mg/ml denatured salmon sperm DNA) and then hybridized at 42°C overnight in fresh  
5 hybridization solution containing 5 ng/ml of the labeled probe. Following hybridization, the filters are washed twice for 10 minutes at room temperature in 2X SSC and 0.1% SDS, and then twice for 30 minutes at 65°C in 0.1X SSC and 0.1% SDS. The blots are then exposed to autoradiography.

10 The expression of GPCR mRNA is localized by *in situ* hybridization. A panel of normal embryonic and adult mouse tissues is fixed in 4% paraformaldehyde, embedded in paraffin, and sectioned at 5  $\mu\text{m}$ . Sectioned tissues are permeabilized in 0.2 M HCl, digested with Proteinase K, and acetylated with triethanolamine and acetic anhydride. Sections are prehybridized for 1 hour at 60°C in hybridization  
15 solution (300 mM NaCl, 20 mM Tris-HCl, pH 8.0, 5 mM EDTA, 1X Denhardt's solution, 0.2% SDS, 10 mM DTT, 0.25 mg/ml tRNA, 25  $\mu\text{g/ml}$  polyA, 25  $\mu\text{g/ml}$  polyC and 50% formamide) and then hybridized overnight at 60°C in the same solution containing 10% dextran and  $2 \times 10^4$  cpm/ $\mu\text{l}$  of a  $^{33}\text{P}$ -labeled antisense riboprobe complementary to the human GPCR gene. The riboprobe is obtained by *in*  
20 *vitro* transcription of a clone containing human GPCR cDNA sequences using standard techniques.

Following hybridization, sections are rinsed in hybridization solution, treated with RNaseA to digest unhybridized probe, and then washed in 0.1X SSC at 55°C for 30 minutes. Sections are then immersed in NTB-2 emulsion (Kodak, Rochester, NY),  
25 exposed for 3 weeks at 4°C, developed, and counterstained with hematoxylin and eosin. Tissue morphology and hybridization signal are simultaneously analyzed by darkfield and standard illumination for brain (one sagittal and two coronal sections), gastrointestinal tract (esophagus, stomach, duodenum, jejunum, ileum, proximal colon, and distal colon), pituitary, liver, lung, heart, spleen, thymus, lymph nodes,  
30 kidney, adrenal, bladder, pancreas, salivary gland, male and female reproductive organs (ovary, oviduct, and uterus in the female; and testis, epididymus, prostate, seminal vesicle, and vas deferens in the male), BAT and WAT (subcutaneous, perirenal), bone (femur), skin, breast, and skeletal muscle.

### Example 5: Production of GPCR Polypeptides

#### A. Expression of GPCR Polypeptides in Bacteria

PCR is used to amplify template DNA sequences encoding a GPCR polypeptide using primers corresponding to the 5' and 3' ends of the sequence. The amplified DNA products may be modified to contain restriction enzyme sites to allow for insertion into expression vectors. PCR products are gel purified and inserted into expression vectors using standard recombinant DNA methodology. An exemplary vector, such as pAMG21 (ATCC no. 98113) containing the lux promoter and a gene encoding kanamycin resistance is digested with Bam HI and Nde I for directional cloning of inserted DNA. The ligated mixture is transformed into an *E. coli* host strain by electroporation and transformants are selected for kanamycin resistance. Plasmid DNA from selected colonies is isolated and subjected to DNA sequencing to confirm the presence of the insert.

Transformed host cells are incubated in 2xYT medium containing 30 µg/mL kanamycin at 30°C prior to induction. Gene expression is induced by the addition of N-(3-oxohexanoyl)-dl-homoserine lactone to a final concentration of 30 ng/mL followed by incubation at either 30°C or 37°C for six hours. The expression of GPCR polypeptide is evaluated by centrifugation of the culture, resuspension and lysis of the bacterial pellets, and analysis of host cell proteins by SDS-polyacrylamide gel electrophoresis.

Inclusion bodies containing GPCR polypeptide are purified as follows. Bacterial cells are pelleted by centrifugation and resuspended in water. The cell suspension is lysed by sonication and pelleted by centrifugation at 195,000 xg for 5 to 10 minutes. The supernatant is discarded, and the pellet is washed and transferred to a homogenizer. The pellet is homogenized in 5 mL of a Percoll solution (75% liquid Percoll and 0.15 M NaCl) until uniformly suspended and then diluted and centrifuged at 21,600 xg for 30 minutes. Gradient fractions containing the inclusion bodies are recovered and pooled. The isolated inclusion bodies are analyzed by SDS-PAGE.

A single band on an SDS polyacrylamide gel corresponding to *E. coli*-produced GPCR polypeptide is excised from the gel, and the N-terminal amino acid sequence is determined essentially as described by Matsudaira *et al.*, 1987, *J. Biol. Chem.* 262:10-35.

#### B. Expression of GPCR Polypeptide in Mammalian Cells

PCR is used to amplify template DNA sequences encoding a GPCR polypeptide using primers corresponding to the 5' and 3' ends of the sequence. The amplified DNA products may be modified to contain restriction enzyme sites to allow for insertion into expression vectors. PCR products are gel purified and inserted into  
5 expression vectors using standard recombinant DNA methodology. An exemplary expression vector, pCEP4 (Invitrogen, Carlsbad, CA), that contains an Epstein-Barr virus origin of replication, may be used for the expression of GPCR polypeptides in 293-EBNA-1 cells. Amplified and gel purified PCR products are ligated into pCEP4 vector and introduced into 293-EBNA cells by lipofection. The transfected cells are  
10 selected in 100 µg/mL hygromycin and the resulting drug-resistant cultures are grown to confluence. The cells are then cultured in serum-free media for 72 hours. The conditioned media is removed and GPCR polypeptide expression is analyzed by SDS-PAGE.

GPCR polypeptide expression may be detected by silver staining.  
15 Alternatively, GPCR polypeptide is produced as a fusion protein with an epitope tag, such as an IgG constant domain or a FLAG epitope, which may be detected by Western blot analysis using antibodies to the peptide tag.

GPCR polypeptides may be excised from an SDS-polyacrylamide gel, or GPCR fusion proteins are purified by affinity chromatography to the epitope tag, and  
20 subjected to N-terminal amino acid sequence analysis as described herein.

#### C. Expression and Purification of GPCR Polypeptide in Mammalian Cells

GPCR polypeptide expression constructs are introduced into 293 EBNA or CHO cells using either a lipofection or calcium phosphate protocol.

25 To conduct functional studies on the GPCR polypeptides that are produced, large quantities of conditioned media are generated from a pool of hygromycin selected 293 EBNA clones. The cells are cultured in 500 cm Nunc Triple Flasks to 80% confluence before switching to serum free media a week prior to harvesting the media. Conditioned media is harvested and frozen at -20°C until purification.

30 Conditioned media is purified by affinity chromatography as described below. The media is thawed and then passed through a 0.2 µm filter. A Protein G column is equilibrated with PBS at pH 7.0, and then loaded with the filtered media. The column is washed with PBS until the absorbance at A<sub>280</sub> reaches a baseline. GPCR polypeptide is eluted from the column with 0.1 M Glycine-HCl at pH 2.7 and

immediately neutralized with 1 M Tris-HCl at pH 8.5. Fractions containing GPCR polypeptide are pooled, dialyzed in PBS, and stored at -70°C.

For Factor Xa cleavage of the human GPCR polypeptide-Fc fusion polypeptide, affinity chromatography-purified protein is dialyzed in 50 mM Tris-HCl, 100 mM NaCl, 2 mM CaCl<sub>2</sub> at pH 8.0. The restriction protease Factor Xa is added to the dialyzed protein at 1/100 (w/w) and the sample digested overnight at room temperature.

#### Example 6: Production of Anti-GPCR Polypeptide Antibodies

Antibodies to GPCR polypeptides may be obtained by immunization with purified protein or with GPCR peptides produced by biological or chemical synthesis. Suitable procedures for generating antibodies include those described in Hudson and Bay, *Practical Immunology* (2nd ed., Blackwell Scientific Publications).

In one procedure for the production of antibodies, animals (typically mice or rabbits) are injected with a GPCR antigen (such as a GPCR polypeptide), and those with sufficient serum titer levels as determined by ELISA are selected for hybridoma production. Spleens of immunized animals are collected and prepared as single cell suspensions from which splenocytes are recovered. The splenocytes are fused to mouse myeloma cells (such as Sp2/0-Ag14 cells), are first incubated in DMEM with 200 U/mL penicillin, 200 µg/mL streptomycin sulfate, and 4 mM glutamine, and are then incubated in HAT selection medium (hypoxanthine, aminopterin, and thymidine). After selection, the tissue culture supernatants are taken from each fusion well and tested for anti-GPCR antibody production by ELISA.

Alternative procedures for obtaining anti-GPCR antibodies may also be employed, such as the immunization of transgenic mice harboring human Ig loci for production of human antibodies, and the screening of synthetic antibody libraries, such as those generated by mutagenesis of an antibody variable domain.

#### Example 7: Expression of GPCR Polypeptide in Transgenic Mice

To assess the biological activity of GPCR polypeptide, a construct encoding a GPCR polypeptide/Fc fusion protein under the control of a liver specific ApoE promoter is prepared. The delivery of this construct is expected to cause pathological changes that are informative as to the function of GPCR polypeptide. Similarly, a construct containing the full-length GPCR polypeptide under the control of the beta



actin promoter is prepared. The delivery of this construct is expected to result in ubiquitous expression.

To generate these constructs, PCR is used to amplify template DNA sequences encoding a GPCR polypeptide using primers that correspond to the 5' and 3' ends of the desired sequence and which incorporate restriction enzyme sites to permit insertion of the amplified product into an expression vector. Following amplification, PCR products are gel purified, digested with the appropriate restriction enzymes, and ligated into an expression vector using standard recombinant DNA techniques. For example, amplified GPCR polypeptide sequences can be cloned into an expression vector under the control of the human  $\beta$ -actin promoter as described by Graham *et al.*, 1997, *Nature Genetics*, 17:272-74 and Ray *et al.*, 1991, *Genes Dev.* 5:2265-73.

Following ligation, reaction mixtures are used to transform an *E. coli* host strain by electroporation and transformants are selected for drug resistance. Plasmid DNA from selected colonies is isolated and subjected to DNA sequencing to confirm the presence of an appropriate insert and absence of mutation. The GPCR polypeptide expression vector is purified through two rounds of CsCl density gradient centrifugation, cleaved with a suitable restriction enzyme, and the linearized fragment containing the GPCR polypeptide transgene is purified by gel electrophoresis. The purified fragment is resuspended in 5 mM Tris, pH 7.4, and 0.2 mM EDTA at a concentration of 2 mg/mL.

Single-cell embryos from BDF1 x BDF1 bred mice are injected as described (International Pub. No. WO 97/23614). Embryos are cultured overnight in a CO<sub>2</sub> incubator and 15-20 two-cell embryos are transferred to the oviducts of a pseudopregnant CD1 female mice. Offspring obtained from the implantation of microinjected embryos are screened by PCR amplification of the integrated transgene in genomic DNA samples as follows. Ear pieces are digested in 20 mL ear buffer (20 mM Tris, pH 8.0, 10 mM EDTA, 0.5% SDS, and 500 mg/mL proteinase K) at 55°C overnight. The sample is then diluted with 200 mL of TE, and 2 mL of the ear sample is used in a PCR reaction using appropriate primers.

At 8 weeks of age, transgenic founder animals and control animals are sacrificed for necropsy and pathological analysis. Portions of spleen are removed and total cellular RNA isolated from the spleens using the Total RNA Extraction Kit (Qiagen) and transgene expression determined by RT-PCR. RNA recovered from spleens is converted to cDNA using the SuperScript<sup>TM</sup> Preamplification System

(Gibco-BRL) as follows. A suitable primer, located in the expression vector sequence and 3' to the GPCR polypeptide transgene, is used to prime cDNA synthesis from the transgene transcripts. Ten mg of total spleen RNA from transgenic founders and controls is incubated with 1 mM of primer for 10 minutes at 70°C and placed on ice.

5 The reaction is then supplemented with 10 mM Tris-HCl, pH 8.3, 50 mM KCl, 2.5 mM MgCl<sub>2</sub>, 10 mM of each dNTP, 0.1 mM DTT, and 200 U of SuperScript II reverse transcriptase. Following incubation for 50 minutes at 42°C, the reaction is stopped by heating for 15 minutes at 72°C and digested with 2U of RNase H for 20 minutes at 37°C. Samples are then amplified by PCR using primers specific for GPCR

10 polypeptide.

#### Example 8: Biological Activity of GPCR Polypeptide in Transgenic Mice

Prior to euthanasia, transgenic animals are weighed, anesthetized by isoflurane and blood drawn by cardiac puncture. The samples are subjected to

15 hematology and serum chemistry analysis. Radiography is performed after terminal exsanguination. Upon gross dissection, major visceral organs are subject to weight analysis.

Following gross dissection, tissues (*i.e.*, liver, spleen, pancreas, stomach, the entire gastrointestinal tract, kidney, reproductive organs, skin and mammary glands,

20 bone, brain, heart, lung, thymus, trachea, esophagus, thyroid, adrenals, urinary bladder, lymph nodes and skeletal muscle) are removed and fixed in 10% buffered Zn-Formalin for histological examination. After fixation, the tissues are processed into paraffin blocks, and 3 mm sections are obtained. All sections are stained with hematoxylin and eosin, and are then subjected to histological analysis.

25 The spleen, lymph node, and Peyer's patches of both the transgenic and the control mice are subjected to immunohistology analysis with B cell and T cell specific antibodies as follows. The formalin fixed paraffin embedded sections are deparaffinized and hydrated in deionized water. The sections are quenched with 3% hydrogen peroxide, blocked with Protein Block (Lipshaw, Pittsburgh, PA), and

30 incubated in rat monoclonal anti-mouse B220 and CD3 (Harlan, Indianapolis, IN). Antibody binding is detected by biotinylated rabbit anti-rat immunoglobulins and peroxidase conjugated streptavidin (BioGenex, San Ramon, CA) with DAB as a chromagen (BioTek, Santa Barbara, CA). Sections are counterstained with hematoxylin.

After necropsy, MLN and sections of spleen and thymus from transgenic animals and control littermates are removed. Single cell suspensions are prepared by gently grinding the tissues with the flat end of a syringe against the bottom of a 100 mm nylon cell strainer (Becton Dickinson, Franklin Lakes, NJ). Cells are washed  
5 twice, counted, and approximately  $1 \times 10^6$  cells from each tissue are then incubated for 10 minutes with 0.5  $\mu$ g CD16/32(Fc $\gamma$ III/II) Fc block in a 20  $\mu$ L volume. Samples are then stained for 30 minutes at 2-8°C in a 100  $\mu$ L volume of PBS (lacking Ca<sup>+</sup> and Mg<sup>+</sup>), 0.1% bovine serum albumin, and 0.01% sodium azide with 0.5  $\mu$ g antibody of  
10 FITC or PE-conjugated monoclonal antibodies against CD90.2 (Thy-1.2), CD45R (B220), CD11b (Mac-1), Gr-1, CD4, or CD8 (PharMingen, San Diego, CA). Following antibody binding, the cells are washed and then analyzed by flow cytometry on a FACScan (Becton Dickinson).

While the present invention has been described in terms of the preferred  
15 embodiments, it is understood that variations and modifications will occur to those skilled in the art. Therefore, it is intended that the appended claims cover all such equivalent variations that come within the scope of the invention as claimed.

## WHAT IS CLAIMED IS:

1. An isolated nucleic acid molecule comprising:
  - (a) the nucleotide sequence as set forth in any of SEQ ID NO: 1, SEQ ID  
5 NO: 3, or SEQ ID NO: 5;
  - (b) a nucleotide sequence encoding the polypeptide as set forth in any of  
SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6;
  - (c) a nucleotide sequence that hybridizes under at least moderately  
stringent conditions to the complement of the nucleotide sequence of either (a) or (b);  
10 or
  - (d) a nucleotide sequence complementary to the nucleotide sequence of  
either (a) or (b).
2. An isolated nucleic acid molecule comprising:
  - 15 (a) a nucleotide sequence encoding a polypeptide that is at least about 70  
percent identical to the polypeptide as set forth in any of SEQ ID NO: 2, SEQ ID NO:  
4, or SEQ ID NO: 6, wherein the encoded polypeptide has an activity of the  
polypeptide set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6;
  - (b) a nucleotide sequence encoding an allelic variant or splice variant of  
20 the nucleotide sequence as set forth in any of SEQ ID NO: 1, SEQ ID NO: 3, or SEQ  
ID NO: 5 or the nucleotide sequence of (a);
  - (c) a region of the nucleotide sequence of any of SEQ ID NO: 1, SEQ ID  
NO: 3, or SEQ ID NO: 5 or the nucleotide sequence of (a) or (b) encoding a  
polypeptide fragment of at least about 25 amino acid residues, wherein the  
25 polypeptide fragment has an activity of the encoded polypeptide as set forth in any of  
SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6, or is antigenic;
  - (d) a region of the nucleotide sequence of any of SEQ ID NO: 1, SEQ ID  
NO: 3, or SEQ ID NO: 5 or the nucleotide sequence of any of (a) - (c) comprising a  
fragment of at least about 16 nucleotides;
  - 30 (e) a nucleotide sequence that hybridizes under at least moderately  
stringent conditions to the complement of the nucleotide sequence of any of (a) - (d);  
or
  - (f) a nucleotide sequence complementary to the nucleotide sequence of  
any of (a) - (d).

3. An isolated nucleic acid molecule comprising:

(a) a nucleotide sequence encoding a polypeptide as set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6 with at least one conservative amino acid substitution, wherein the encoded polypeptide has an activity of the polypeptide set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6;

(b) a nucleotide sequence encoding a polypeptide as set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6 with at least one amino acid insertion, wherein the encoded polypeptide has an activity of the polypeptide set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6;

(c) a nucleotide sequence encoding a polypeptide as set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6 with at least one amino acid deletion, wherein the encoded polypeptide has an activity of the polypeptide set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6;

(d) a nucleotide sequence encoding a polypeptide as set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6 that has a C- and/or N- terminal truncation, wherein the encoded polypeptide has an activity of the polypeptide set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6;

(e) a nucleotide sequence encoding a polypeptide as set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6 with at least one modification that is an amino acid substitution, an amino acid insertion, an amino acid deletion, C-terminal truncation, or N-terminal truncation, wherein the encoded polypeptide has an activity of the polypeptide set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6;

(f) a nucleotide sequence of any of (a) - (e) comprising a fragment of at least about 16 nucleotides;

(g) a nucleotide sequence that hybridizes under at least moderately stringent conditions to the complement of the nucleotide sequence of any of (a) - (f);  
or

(h) a nucleotide sequence complementary to the nucleotide sequence of any of (a) - (e).

4. A vector comprising the nucleic acid molecule of any of Claims 1, 2, or 3.

5. A host cell comprising the vector of Claim 4.
6. The host cell of Claim 5 that is a eukaryotic cell.
- 5 7. The host cell of Claim 5 that is a prokaryotic cell.
8. A process of producing a GPCR polypeptide comprising culturing the host cell of Claim 5 under suitable conditions to express the polypeptide, and  
10 optionally isolating the polypeptide from the culture.
9. A polypeptide produced by the process of Claim 8.
10. The process of Claim 8, wherein the nucleic acid molecule comprises  
15 promoter DNA other than the promoter DNA for the native GPCR polypeptide operatively linked to the DNA encoding the GPCR polypeptide.
11. The isolated nucleic acid molecule according to Claim 2, wherein the percent identity is determined using a computer program selected from the group  
20 consisting of GAP, BLASTN, FASTA, BLASTA, BLASTX, BestFit, and the Smith-Waterman algorithm.
12. A process for determining whether a compound inhibits GPCR polypeptide activity or GPCR polypeptide production comprising exposing a cell  
25 according to any of Claims 5, 6, or 7 to the compound and measuring GPCR polypeptide activity or GPCR polypeptide production in said cell.
13. An isolated polypeptide comprising the amino acid sequence as set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6.  
30
14. An isolated polypeptide comprising:
- (a) an amino acid sequence for an ortholog of any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6;

- (b) an amino acid sequence that is at least about 70 percent identical to the amino acid sequence of any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6, wherein the polypeptide has an activity of the polypeptide set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6;
- 5 (c) a fragment of the amino acid sequence set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6 comprising at least about 25 amino acid residues, wherein the fragment has an activity of the polypeptide set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6, or is antigenic; or
- 10 (d) an amino acid sequence for an allelic variant or splice variant of the amino acid sequence as set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6 or the amino acid sequence of either (a) or (b).
15. An isolated polypeptide comprising:
- (a) the amino acid sequence as set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6 with at least one conservative amino acid substitution, wherein the polypeptide has an activity of the polypeptide set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6;
- 20 (b) the amino acid sequence as set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6 with at least one amino acid insertion, wherein the polypeptide has an activity of the polypeptide set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6;
- (c) the amino acid sequence as set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6 with at least one amino acid deletion, wherein the polypeptide has an activity of the polypeptide set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6;
- 25 (d) the amino acid sequence as set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6 that has a C- and/or N- terminal truncation, wherein the polypeptide has an activity of the polypeptide set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6; or
- 30 (e) the amino acid sequence as set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6 with at least one modification that is an amino acid substitution, an amino acid insertion, an amino acid deletion, C-terminal truncation, or N-terminal truncation, wherein the polypeptide has an activity of the polypeptide set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6.

16. An isolated polypeptide encoded by the nucleic acid molecule of any of Claims 1, 2, or 3, wherein the polypeptide has an activity of the polypeptide set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6.

5

17. The isolated polypeptide according to Claim 14, wherein the percent identity is determined using a computer program selected from the group consisting of GAP, BLASTP, FASTA, BLASTA, BLASTX, BestFit, and the Smith-Waterman algorithm.

10

18. A selective binding agent or fragment thereof that specifically binds the polypeptide of any of Claims 13, 14, or 15.

19. The selective binding agent or fragment thereof of Claim 18 that specifically binds the polypeptide comprising the amino acid sequence as set forth in any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6, or a fragment thereof.

15

20. The selective binding agent of Claim 18 that is an antibody or fragment thereof.

20

21. The selective binding agent of Claim 18 that is a humanized antibody.

22. The selective binding agent of Claim 18 that is a human antibody or fragment thereof.

25

23. The selective binding agent of Claim 18 that is a polyclonal antibody or fragment thereof.

24. The selective binding agent Claim 18 that is a monoclonal antibody or fragment thereof.

30

25. The selective binding agent of Claim 18 that is a chimeric antibody or fragment thereof.



26. The selective binding agent of Claim 18 that is a CDR-grafted antibody or fragment thereof.

27. The selective binding agent of Claim 18 that is an antiidiotypic  
5 antibody or fragment thereof.

28. The selective binding agent of Claim 18 that is a variable region fragment.

10 29. The variable region fragment of Claim 28 that is a Fab or a Fab' fragment.

30. A selective binding agent or fragment thereof comprising at least one complementarity determining region with specificity for a polypeptide having the  
15 amino acid sequence of any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6.

31. The selective binding agent of Claim 18 that is bound to a detectable label.

20 32. The selective binding agent of Claim 18 that antagonizes GPCR polypeptide biological activity.

33. A method for treating, preventing, or ameliorating a GPCR polypeptide-related disease, condition, or disorder comprising administering to a  
25 patient an effective amount of a selective binding agent according to Claim 18.

34. A selective binding agent produced by immunizing an animal with a polypeptide comprising an amino acid sequence of any of SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6,

30

35. A hybridoma that produces a selective binding agent capable of binding a polypeptide according to any of Claims 13, 14, or 15.

36. A method of detecting or quantitating the amount of GPCR polypeptide using the anti-GPCR antibody or fragment of Claim 18.

37. A kit for detecting or quantitating the amount of GPCR polypeptide in  
5 a biological sample, comprising the selective binding agent of Claim 18.

38. A composition comprising the polypeptide of any of Claims 13, 14, or 15, and a pharmaceutically acceptable formulation agent.

10 39. The composition of Claim 38, wherein the pharmaceutically acceptable formulation agent is a carrier, adjuvant, solubilizer, stabilizer, or anti-oxidant.

40. A polypeptide comprising a derivative of the polypeptide of any of  
15 Claims 13, 14, or 15.

41. The polypeptide of Claim 40 that is covalently modified with a water-soluble polymer.

20 42. The polypeptide of Claim 41, wherein the water-soluble polymer is polyethylene glycol, monomethoxy-polyethylene glycol, dextran, cellulose, poly-(N-vinyl pyrrolidone) polyethylene glycol, propylene glycol homopolymers, polypropylene oxide/ethylene oxide co-polymers, polyoxyethylated polyols, or polyvinyl alcohol.

25 43. A composition comprising a nucleic acid molecule of any of Claims 1, 2, or 3 and a pharmaceutically acceptable formulation agent.

44. The composition of Claim 43, wherein said nucleic acid molecule is  
30 contained in a viral vector.

45. A viral vector comprising a nucleic acid molecule of any of Claims 1, 2, or 3.

46. A fusion polypeptide comprising the polypeptide of any of Claims 13, 14, or 15 fused to a heterologous amino acid sequence.

47. The fusion polypeptide of Claim 46, wherein the heterologous amino acid sequence is an IgG constant domain or fragment thereof.

48. A method for treating, preventing, or ameliorating a medical condition comprising administering to a patient the polypeptide of any of Claims 13, 14, or 15, or the polypeptide encoded by the nucleic acid of any of Claims 1, 2, or 3.

10

49. A method of diagnosing a pathological condition or a susceptibility to a pathological condition in a subject comprising:

(a) determining the presence or amount of expression of the polypeptide of any of Claims 13, 14, or 15, or the polypeptide encoded by the nucleic acid molecule of any of Claims 1, 2, or 3 in a sample; and

15

(b) diagnosing a pathological condition or a susceptibility to a pathological condition based on the presence or amount of expression of the polypeptide.

20

50. A device, comprising:

(a) a membrane suitable for implantation; and

(b) cells encapsulated within said membrane, wherein said cells secrete a protein of any of Claims 13, 14, or 15; and

said membrane is permeable to said protein and impermeable to materials detrimental to said cells.

25

51. A method of identifying a compound that binds to a GPCR polypeptide comprising:

(a) contacting the polypeptide of any of Claims 13, 14, or 15 with a compound; and

30

(b) determining the extent of binding of the GPCR polypeptide to the compound.

52. The method of Claim 51, further comprising determining the activity of the polypeptide when bound to the compound.

53. A method of modulating levels of a polypeptide in an animal  
5 comprising administering to the animal the nucleic acid molecule of any of Claims 1, 2, or 3.

54. A transgenic non-human mammal comprising the nucleic acid molecule of any of Claims 1, 2, or 3.  
10

55. A process for determining whether a compound inhibits GPCR polypeptide activity or GPCR polypeptide production comprising exposing a transgenic mammal according to Claim 54 to the compound, and measuring GPCR polypeptide activity or GPCR polypeptide production in said mammal.  
15

56. A nucleic acid molecule of any of Claims 1, 2, or 3 attached to a solid support.

57. An array of nucleic acid molecules comprising at least one nucleic acid  
20 molecule of any of Claims 1, 2, or 3.

58. An isolated polypeptide comprising the amino acid sequence as set forth in SEQ ID NO: 2 with at least one conservative amino acid substitution that is a aspartic acid at position 2; leucine at position 8; glutamic acid at position 12; proline at position 13; threonine at position 15; leucine at position 24; valine at position 25; leucine at position 27; isoleucine or leucine at position 35; serine at position 49; isoleucine at position 52; isoleucine at position 82; leucine at position 83; alanine at position 87; leucine at position 90; valine at position 91; lysine or methionine at position 94; valine at position 111; methionine at position 123; alanine at position 126; N at position 129; threonine at position 131; alanine at position 134; threonine at position 135; alanine at position 136; valine at position 138; threonine at position 141; methionine at position 152; serine at position 154; arginine at position 159; glycine at position 160; methionine at position 161; leucine or valine at position 162; serine at position 163; valine at position 179; leucine at position 187; threonine at position 190;  
30

valine at position 197; asparagine at position 198; valine at position 199; glutamine at position 205; threonine at position 210; arginine at position 216; arginine at position 217; serine at position 227; leucine at position 238; threonine at position 249; threonine or isoleucine at position 258; valine at position 262; leucine at position 266;  
5 leucine at position 269; leucine at position 285; alanine at position 290; threonine at position 293; arginine at position 295; arginine at position 300; arginine at position 301; arginine at position 304; threonine at position 305; glutamine at position 306; alanine at position 307; arginine at position 308; serine at position 310; glycine at position 319; serine at position 320; lysine at position 321; serine at position 322;  
10 threonine at position 324; aspartic acid at position 325; glycine at position 326; valine at position 327; arginine at position 329; serine at position 330; arginine at position 332; proline at position 334; glycine at position 339; leucine at position 340; glutamine at position 341; and valine at position 342; wherein the polypeptide has an activity of the polypeptide set forth in SEQ ID NO: 2.

15

FIG. 1A

atg tac aac ggg tcg tgc tgc cgc atc gag ggg gac acc atc tcc cag	48
Met Tyr Asn Gly Ser Cys Cys Arg Ile Glu Gly Asp Thr Ile Ser Gln	
1 5 10 15	
gtg atg ccg ccg ctg ctc att gtg gcc ttt gtg ctg ggc gca cta ggc	96
Val Met Pro Pro Leu Leu Ile Val Ala Phe Val Leu Gly Ala Leu Gly	
20 25 30	
aat ggg gtc gcc ctg tgt ggt ttc tgc ttc cac atg aag acc tgg aag	144
Asn Gly Val Ala Leu Cys Gly Phe Cys Phe His Met Lys Thr Trp Lys	
35 40 45	
ccc agc act gtt tac ctt ttc aat ttg gcc gtg gct gat ttc ctc ctt	192
Pro Ser Thr Val Tyr Leu Phe Asn Leu Ala Val Ala Asp Phe Leu Leu	
50 55 60	
atg atc tgc ctg cct ttt cgg aca gac tat tac ctc aga cgt aga cac	240
Met Ile Cys Leu Pro Phe Arg Thr Asp Tyr Tyr Leu Arg Arg Arg His	
65 70 75 80	
tgg gct ttt ggg gac att ccc tgc cga gtg ggg ctc ttc acg ttg gcc	288
Trp Ala Phe Gly Asp Ile Pro Cys Arg Val Gly Leu Phe Thr Leu Ala	
85 90 95	
atg aac agg gcc ggg agc atc gtg ttc ctt acg gtg gtg gct gcg gac	336
Met Asn Arg Ala Gly Ser Ile Val Phe Leu Thr Val Val Ala Ala Asp	
100 105 110	
agg tat ttc aaa gtg gtc cac ccc cac cac gcg gtg aac act atc tcc	384
Arg Tyr Phe Lys Val Val His Pro His His Ala Val Asn Thr Ile Ser	
115 120 125	
acc cgg gtg gcg gct ggc atc gtc tgc acc ctg tgg gcc ctg gtc atc	432
Thr Arg Val Ala Ala Gly Ile Val Cys Thr Leu Trp Ala Leu Val Ile	
130 135 140	
ctg gga aca gtg tat ctt ttg ctg gag aac cat ctc tgc gtg caa gag	480
Leu Gly Thr Val Tyr Leu Leu Leu Glu Asn His Leu Cys Val Gln Glu	
145 150 155 160	
acg gcc gtc tcc tgt gag agc ttc atc atg gag tcg gcc aat ggc tgg	528
Thr Ala Val Ser Cys Glu Ser Phe Ile Met Glu Ser Ala Asn Gly Trp	
165 170 175	
cat gac atc atg ttc cag ctg gag ttc ttt atg ccc ctc ggc atc atc	576
His Asp Ile Met Phe Gln Leu Glu Phe Phe Met Pro Leu Gly Ile Ile	
180 185 190	
tta ttt tgc tcc ttc aag att gtt tgg agc ctg agg cgg agg cag cag	624
Leu Phe Cys Ser Phe Lys Ile Val Trp Ser Leu Arg Arg Arg Gln Gln	
195 200 205	

## FIG. 1B

ctg gcc aga cag gct cgg atg aag aag gcg acc cgg ttc atc atg gtg	672
Leu Ala Arg Gln Ala Arg Met Lys Lys Ala Thr Arg Phe Ile Met Val	
210 215 220	
gtg gca att gtg ttc atc aca tgc tac ctg ccc agc gtg tct gct aga	720
Val Ala Ile Val Phe Ile Thr Cys Tyr Leu Pro Ser Val Ser Ala Arg	
225 230 235 240	
ctc tat ttc ctc tgg acg gtg ccc tgg agt gcc tgc gat ccc tct gtc	768
Leu Tyr Phe Leu Trp Thr Val Pro Ser Ser Ala Cys Asp Pro Ser Val	
245 250 255	
cat ggg gcc ctg cac ata acc ctc agc ttc acc tac atg aac agc atg	816
His Gly Ala Leu His Ile Thr Leu Ser Phe Thr Tyr Met Asn Ser Met	
260 265 270	
ctg gat ccc ctg gtg tat tat ttt tca agc ccc tcc ttt ccc aaa ttc	864
Leu Asp Pro Leu Val Tyr Tyr Phe Ser Ser Pro Ser Phe Pro Lys Phe	
275 280 285	
tac aac aag ctc aaa atc tgc agt ctg aaa ccc aag cag cca gga cac	912
Tyr Asn Lys Leu Lys Ile Cys Ser Leu Lys Pro Lys Gln Pro Gly His	
290 295 300	
tca aaa aca caa agg ccg gaa gag atg cca att tcg aac ctc ggt cgc	960
Ser Lys Thr Gln Arg Pro Glu Glu Met Pro Ile Ser Asn Leu Gly Arg	
305 310 315 320	
agg agt tgc atc agt gtg gca aat agt ttc caa agc cag tct gat ggg	1008
Arg Ser Cys Ile Ser Val Ala Asn Ser Phe Gln Ser Gln Ser Asp Gly	
325 330 335	
caa tgg gat ccc cac att gtt gag tgg cac	1038
Gln Trp Asp Pro His Ile Val Glu Trp His	
340 345	

## FIG. 2A

```

gaaaaagaca aaaccagaaa aagaaaagtc atctccaggg ctcgatctag caacgagtct 60
gtagcatgta tagcgtcgga cccccgagct gcaaccaga aatgtacact cgtgggaaac 120
cgcttgacac ccagagcctg acccagctgc aggcctcaac tctgtagggg acgtgcagct 180
cgtgatccaa gcctaggaga aaggacttgc tgccggcttt catttcttg ctgaagtttc 240
tctcgtgggt gcagcgctg catcccaggg tgatgagggt aggggccag ctgctagagg 300
agccctagtg ttcggatagg cagctgtgcc tctgtgccgg ccaccttg atg cca gtc 358
                                     Met Pro Val
                                     1

ctc tct cca act gct atg gac aac ggg tcg tgc tgt ctc atc gag ggg 406
Leu Ser Pro Thr Ala Met Asp Asn Gly Ser Cys Cys Leu Ile Glu Gly
      5                      10                      15

gag ccc atc tcc cag gtg atg cct cct cta ctc atc ctg gtc ttc gtg 454
Glu Pro Ile Ser Gln Val Met Pro Pro Leu Leu Ile Leu Val Phe Val
      20                      25                      30                      35

ctt ggc gcc ctg ggc aac ggc ata gcc ctg tgc ggc ttc tgc ttt cac 502
Leu Gly Ala Leu Gly Asn Gly Ile Ala Leu Cys Gly Phe Cys Phe His
                      40                      45                      50

atg aag acc tgg aag tca agc act att tac ctt ttc aac ttg gct gtg 550
Met Lys Thr Trp Lys Ser Ser Thr Ile Tyr Leu Phe Asn Leu Ala Val
                      55                      60                      65

gcc gat ttt ctc ctc atg atc tgc tta ccc ctt cgg aca gac tac tac 598
Ala Asp Phe Leu Leu Met Ile Cys Leu Pro Leu Arg Thr Asp Tyr Tyr
                      70                      75                      80

ctc aga cgc aga cac tgg att ttt gga gat atc gcc tgt cgc ctg gtc 646
Leu Arg Arg Arg His Trp Ile Phe Gly Asp Ile Ala Cys Arg Leu Val
                      85                      90                      95

ctc ttc aag ctg gcc atg aat agg gcc ggg agc att gtc ttc ctc act 694
Leu Phe Lys Leu Ala Met Asn Arg Ala Gly Ser Ile Val Phe Leu Thr
100                      105                      110                      115

gtg gtg gct gtg gat agg tat ttc aaa gtg gtc cac ccc cac cat atg 742
Val Val Ala Val Asp Arg Tyr Phe Lys Val Val His Pro His His Met
                      120                      125                      130

gtg aat gcc atc tcc aac cgg act gcc gcc gcc acc gcc tgt gtc ctc 790
Val Asn Ala Ile Ser Asn Arg Thr Ala Ala Ala Thr Ala Cys Val Leu
                      135                      140                      145

tgg act ttg gtc atc ttg ggg act gtg tat ctt ctg atg gag agt cac 838
Trp Thr Leu Val Ile Leu Gly Thr Val Tyr Leu Leu Met Glu Ser His
                      150                      155                      160

```



## FIG. 2B

ctg tgt gtg cag ggg aca ctg tgc tcc tgt gag agc ttc atc atg gag	886
Leu Cys Val Gln Gly Thr Leu Ser Ser Cys Glu Ser Phe Ile Met Glu	
165 170 175	
tca gcc aac ggg tgg cac gat gtc atg ttc cag ctg gag ttc ttc ctg	934
Ser Ala Asn Gly Trp His Asp Val Met Phe Gln Leu Glu Phe Phe Leu	
180 185 190 195	
ccc ctg aca atc atc ttg ttc tgc tgc gtc aac gtt gtt tgg agc ctg	982
Pro Leu Thr Ile Ile Leu Phe Cys Ser Val Asn Val Val Trp Ser Leu	
200 205 210	
aga cgg agg cag cag ctg acc aga cag gct cgg atg agg agg gcc acc	1030
Arg Arg Arg Gln Gln Leu Thr Arg Gln Ala Arg Met Arg Arg Ala Thr	
215 220 225	
cgg ttc atc atg gtg gtg gct tct gtg ttc atc acg tgt tac ctg ccc	1078
Arg Phe Ile Met Val Val Ala Ser Val Phe Ile Thr Cys Tyr Leu Pro	
230 235 240	
agc gtg ctg gct agg ctc tac ttc ctc tgg acg gtg ccc act agt gcc	1126
Ser Val Leu Ala Arg Leu Tyr Phe Leu Trp Thr Val Pro Thr Ser Ala	
245 250 255	
tgt gac ccc tct gtc cac aca gcc ctc cac gtc acc ctg agc ttc acc	1174
Cys Asp Pro Ser Val His Thr Ala Leu His Val Thr Leu Ser Phe Thr	
260 265 270 275	
tac ctg aac agt atg ctg gat ccc ctt gta tat tac ttc tca agc ccc	1222
Tyr Leu Asn Ser Met Leu Asp Pro Leu Val Tyr Tyr Phe Ser Ser Pro	
280 285 290	
tgc ctc ccc aaa ttc tac gcc aag ctc aca atc tgc agc ctg aag ccc	1270
Ser Leu Pro Lys Phe Tyr Ala Lys Leu Thr Ile Cys Ser Leu Lys Pro	
295 300 305	
aaa cgc cca gga cgc acg aag acg cgg agg tca gaa gag atg cca att	1318
Lys Arg Pro Gly Arg Thr Lys Thr Arg Arg Ser Glu Glu Met Pro Ile	
310 315 320	
tgc aac ctc tgc agt aag agc tcc atc gat ggg gca aat cgt tcc cag	1366
Ser Asn Leu Cys Ser Lys Ser Ser Ile Asp Gly Ala Asn Arg Ser Gln	
325 330 335	
agg cca tct gac ggg cag tgg gat ctc caa gtg tgt tgaatgccat	1412
Arg Pro Ser Asp Gly Gln Trp Asp Leu Gln Val Cys	
340 345 350	
taagacaaac agcccaacaa cgaggcagag aaatgggcaa tgtgagttaa atctgaagg	1472
tggaggactt gaagatgtcc cctcccactc ttagctgtat ctttctcact caggtagaaa	1532
tgggatccac cctgcttgac cttttccaga aggttccaaa ccggttggtt gtgtttaaat	1592

## FIG. 2C

actctgatag caatggtgaa ggggcagcgt gtgagtgtga aggaaaccgt ggggtgtcggg 1652  
ttaggaacta cctggagccc gtgtcgcttt gcatggctga gaaaagcggg atgagcctgg 1712  
ctgggtcttg ttctagctcg gagagagtta acgatctcaa taactcgtcg gtatttcctg 1772  
gactgaaaaa aatagaaact gcaactgagtc aatacactta tttccagctg agcgagaccc 1832  
tttactgcag gacacccgga cctagccgtt tttttaaatc ttccctgggg agcctccaca 1892  
catttcaagg tttgaacatc cagggtggccc aggagggcag caaaaagaat ctattctaac 1952  
cttgctggcg ccacgatatt ttgctgattt taagtggttt catcctttgt ttttcttttg 2012  
tttgcatttc aaagaagatg ctgagggact tgtccacctg atatcagcta tcgtttctcc 2072  
agtgggaact gagagcctgt ttacggcagc aatggtgggg ggggggtgct tcctggtatt 2132  
tgggatgggt taattcaagc atggttgctt ttactgctt aatgcatgaa tttgagctga 2192  
aatcctccct tctcaagtct ttgtttaatc cacagtatgt tgtccacccc tgtccagcat 2252  
cctcgtctgt tttgtctttg gtgctgggca ttgaacttgg gaccttattc gagccaggca 2312  
agcactgacc actgaactgc actcccaaac cccttggtgc ccttttagct gtagcgttgt 2372  
tagccaactt ttgggagaaa gcaaagcact agaggtggca gcaacagttt agctcaatgt 2432  
cctttcgtca gtgtctagac ttctggtcag ccatccgggt ctcctattgg gggcctccct 2492  
caagcacata tgttctccca aatactaccc agaattctca cagctagggtg attctgtgaa 2552  
agtccaggct gccctgtcc tggagaagga gaaatagaat ccgtgttaac cttagtccca 2612  
ctttcaagcc acaaaagtgg tgacagccat tcaactcttc cagttcccag ggtactctcc 2672  
ccagggaagg gaccttgaca tttatgtcta aagacataaa ttagatgctc ctcaaggttg 2732  
tccctgtggc ttcctttgcc agaggtgttg aagcctagggt gcgaaaatca gtctgactgc 2792  
agggctggtg agaaggctca gtaggtaaac aggtctgccg ccaagccacc aacctgagta 2852  
tccctgagac ccacgtggta ggagaacacc aaatcccaaa ggtggttctg tcctccacct 2912  
atgttcacgc atgcgcacat gcacgcatgc gcacgcgcgc acaccaccc accaaatata 2972  
ttaatgcaat aaaaatttta ttggctacac ggtcaagttt gaatcttagt ttaaagtctt 3032  
attagacatg tgctcgtagg gaagacttta ttttaaccca ctcagttttg atgttcagca 3092  
gggttaatac tgatgccaaa gggtttggga gcaaattcaa tgactgtaca ctcataagca 3152  
tgagaaatct gttgttccca ggtctgcccg gaagaagacc atgtgcgtgt agtagttgat 3212

FIG. 2D

aaataaatag ttgctgaaca actataatcg ctccaaaaa

3251

## FIG. 3A

tggtagcgct gcaggtaccg gtccggaatt cccgggtcga cccacgcgtc cgcaaaacta 60  
 gaaaagcaaa atcgccccca ggggtggacc cagcgacaag tctgctgctg ggctggcatc 120  
 agacccccaa gctgcagcct ggcaatgtac gcttttggaa aactgctctc gcctcagagc 180  
 ctgaccagc tgcaggcttc acctctgtag gggacatgca gcttgtgatc caggctgagg 240  
 agaaaggacc tgctgtcggc ttccatttcc tgactgaagt tgggtgatga ggtaggggc 300  
 ccagctgcc aaggggaacca tagtgttcag ataggcagct gtgcctttgt gtcagccacc 360  
 tagg atg ctc ttc ctc tct ccg agt gct atg gac aac ggg tcg tgc tgt 409  
 Met Leu Phe Leu Ser Pro Ser Ala Met Asp Asn Gly Ser Cys Cys  
 1 5 10 15  
 ctc atc gag ggg gaa ccc atc acc cag gta atg cca cct tta ctc atc 457  
 Leu Ile Glu Gly Glu Pro Ile Thr Gln Val Met Pro Pro Leu Leu Ile  
 20 25 30  
 ctg gcc ttc ctg ctt gga gcc ctg ggc aac ggc cta gcc ctg tgt ggt 505  
 Leu Ala Phe Leu Leu Gly Ala Leu Gly Asn Gly Leu Ala Leu Cys Gly  
 35 40 45  
 ttc tgc ttt cac atg aag acc tgg aag tcg agc act att tac ctt ttc 553  
 Phe Cys Phe His Met Lys Thr Trp Lys Ser Ser Thr Ile Tyr Leu Phe  
 50 55 60  
 aac ttg gct gta gcc gat ttt ctc ctc atg atc tgc cta ccc ctt cgg 601  
 Asn Leu Ala Val Ala Asp Phe Leu Leu Met Ile Cys Leu Pro Leu Arg  
 65 70 75  
 aca gac tac tac ctc aga cgt agg cat tgg att ttg ggg gat att ccc 649  
 Thr Asp Tyr Tyr Leu Arg Arg Arg His Trp Ile Leu Gly Asp Ile Pro  
 80 85 90 95  
 tgc cgc ctg gtc ctc ttc atg ctg gcc atg aat agg gcc gga agc att 697  
 Cys Arg Leu Val Leu Phe Met Leu Ala Met Asn Arg Ala Gly Ser Ile  
 100 105 110  
 gtc ttc ctc act gtg gtg gcc gtg gac agg tat ttc aaa gtg gtc cac 745  
 Val Phe Leu Thr Val Val Ala Val Asp Arg Tyr Phe Lys Val Val His  
 115 120 125  
 ccc cac cat atg gtg aac gcc atc tcc aat cgg act gca gct gcc atc 793  
 Pro His His Met Val Asn Ala Ile Ser Asn Arg Thr Ala Ala Ala Ile  
 130 135 140  
 gtc tgt gtc ctc tgg act ttg gtc atc ttg ggg act gtg tat ctt ctg 841  
 Val Cys Val Leu Trp Thr Leu Val Ile Leu Gly Thr Val Tyr Leu Leu  
 145 150 155

## FIG. 3B

atg gag agt cac ctg tgt gtg cgg ggg atg gtg tca tct tgt gag agc	889
Met Glu Ser His Leu Cys Val Arg Gly Met Val Ser Ser Cys Glu Ser	
160 165 170 175	
ttc atc atg gag tca gcc aac ggg tgg cac gat atc atg ttc cag ctg	937
Phe Ile Met Glu Ser Ala Asn Gly Trp His Asp Ile Met Phe Gln Leu	
180 185 190	
gag ttc ttc ctg ccc ctg acc atc atc ttg ttc tgc tcc ttc aaa gtt	985
Glu Phe Phe Leu Pro Leu Thr Ile Ile Leu Phe Cys Ser Phe Lys Val	
195 200 205	
gtt tgg agc ctg aga cag agg caa cag ctg acc aga cag gct cgg atg	1033
Val Trp Ser Leu Arg Gln Arg Gln Gln Leu Thr Arg Gln Ala Arg Met	
210 215 220	
agg agg gcc acc cgg ttc atc atg gtg gtg gct tcc gtg ttc atc acg	1081
Arg Arg Ala Thr Arg Phe Ile Met Val Val Ala Ser Val Phe Ile Thr	
225 230 235	
tgt tac ctg ccc agc gtg ttg gcg agg ctc tac ttc ctc tgg acg gtg	1129
Cys Tyr Leu Pro Ser Val Leu Ala Arg Leu Tyr Phe Leu Trp Thr Val	
240 245 250 255	
ccc tcc agt gct tgt gac ccc tct gtc cac ata gct ctc cat gtc acc	1177
Pro Ser Ser Ala Cys Asp Pro Ser Val His Ile Ala Leu His Val Thr	
260 265 270	
ctg agt ctc acc tac ctg aac agc atg ctg gac cct ctt gtg tac tac	1225
Leu Ser Leu Thr Tyr Leu Asn Ser Met Leu Asp Pro Leu Val Tyr Tyr	
275 280 285	
ttt tca agc ccc tcg ttc ccc aaa ttc tac gcc aag ctc aaa atc cgc	1273
Phe Ser Ser Pro Ser Phe Pro Lys Phe Tyr Ala Lys Leu Lys Ile Arg	
290 295 300	
agc ttg aaa ccc aga cgc cca gga cgc tcg cag gca cgg agg tcg gaa	1321
Ser Leu Lys Pro Arg Arg Pro Gly Arg Ser Gln Ala Arg Arg Ser Glu	
305 310 315	
gag atg cca att tcg aat ctc tgt cgt aag agt tcc acc gat gtg gta	1369
Glu Met Pro Ile Ser Asn Leu Cys Arg Lys Ser Ser Thr Asp Val Val	
320 325 330 335	
aat agt tcc cag agg ccg tct gac ggg cag tgg ggt ctc caa gtg tgt	1417
Asn Ser Ser Gln Arg Pro Ser Asp Gly Gln Trp Gly Leu Gln Val Cys	
340 345 350	
tgaatgccat gaagacaaat ggcccagcag caaagcagag acctgggcaa ctgtgagtta	1477
aatctgaagg gtgagggact tgaaaaatga cagccccccc ccccgccca cccgcccgcc	1537
cgccccgctc tttctcagct gtgtctttct cactcaagta gaagcaaaat ctaaaaaaaaa	1597

# FIG. 3C

aaaaaaaaa aaaaaaaaaa agggcggccg ctctagagga tccaagctta cgtacgcgtg 1657  
catgcgacgt c 1668

FIG. 4

TM1

hu_GPCR	1	MYNGSCCRIEGDTISQVMPPLLIVAFVLGALGNGVALCGFCF	42
mu_GPCR	1	MPVLSPTAMDNGSCCLIEGEPISQVMPPLLILVFVLGALGNGIALCGFCF	50
ra_GPCR	1	MLFLSPSAMDNGSCCLIEGEPITQVMPPLLILAFLLGALGNGLALCGFCF	50
		* * * * *	

TM2                      extracellular loop 1

hu_GPCR	43	HMKTWKPSTVYLFNLAVADFLLMICLPFRTDYYLRRRHWAFGDIPCRVGL	92
mu_GPCR	51	HMKTWKSSTIYLFNLAVADFLLMICLPFRTDYYLRRRHWFPGDIACRLVL	100
ra_GPCR	51	HMKTWKSSTIYLFNLAVADFLLMICLPFRTDYYLRRRHWFILGDIPCRVL	100
		*****	

TM3                      TM4

hu_GPCR	93	FTLAMNRAGSIVFLTVVAADRYFKVPHHVAANTISTRVAAGIVCTLWAL	142
mu_GPCR	101	FKLAMNRAGSIVFLTVVAVDRYFKVPHHVMNAISNRTAAATACVLWTL	150
ra_GPCR	101	FMLAMNRAGSIVFLTVVAVDRYFKVPHHVMNAISNRTAAATVCVLWTL	150
		*****	

extracellular loop 2                      TM5

hu_GPCR	143	VILGTVYLLLENHLCVQETAVSCSFIMESANGWHDIMFOLEFFMPLGII	192
mu_GPCR	151	VILGTVYLLMESHLCVQGTLSCEFSIMESANGWHDVMFOLEFFPLPTII	200
ra_GPCR	151	VILGTVYLLMESHLCVRGMVSSCEFSIMESANGWHDIMFOLEFFPLPTII	200
		*****	

TM6

hu_GPCR	193	LFCSFKIVWSLRRRQQLARQARMKKATRFIMVVAIVFITCYLPSVSARLY	242
mu_GPCR	201	LFCSVNVVWSLRRRQQLTRQARMRRATRFIMVVASVFITCYLPSVLARLY	250
ra_GPCR	201	LFCSFKVWSLRRRQQLTRQARMRRATRFIMVVASVFITCYLPSVLARLY	250
		****	

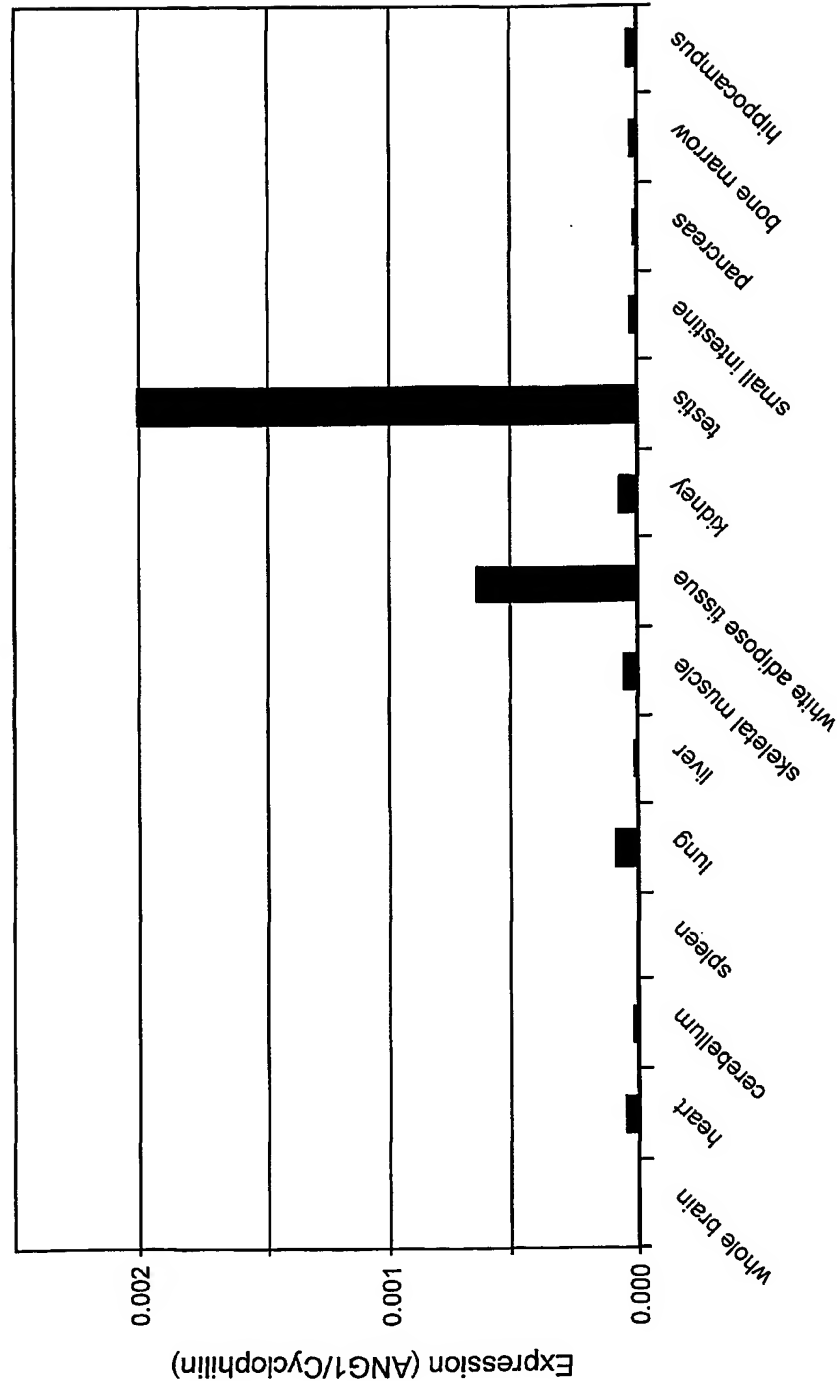
extracellular loop 3                      TM7

hu_GPCR	243	FLWTVPSACDPSVHGALHITLSFTYMNSMLDPLVYFSSPSFPKFKYNKL	292
mu_GPCR	251	FLWTVPTSACDPSVHTALHVTLSFTYLNSMLDPLVYFSSPSLPKFKYAKL	300
ra_GPCR	251	FLWTVPSACDPSVHTALHVTLSFTYLNSMLDPLVYFSSPSFPKFKYAKL	300
		*****	

hu_GPCR	293	KICSLKPKQPGHKTQRPEEMPISNLGRRSCISVANSFQSQSDGQWDPHI	342
mu_GPCR	301	TICSLKPKRPGRTKTRSEEMPISNLCSKSSIDGANRSQRPSDGQWDLQV	350
ra_GPCR	301	KIRSLKPRRPGRSQARRSEEMPISNLCKRSSTDVNSSQRPSDGQWGLQV	350
		* * * * *	

hu_GPCR	343	VEWH	346
mu_GPCR	351	C	351
ra_GPCR	351	C	351

FIG. 5





## SEQUENCE LISTING

<110> Elliott, Steven G.  
 Rogers, Norma  
 Busse, Leigh Anne

<120> G-Protein Coupled Receptor Molecules and Uses Thereof

<130> 02-076-A

<140>

<141>

<150> 60/269,040

<151> 2001-02-14

<160> 22

<170> PatentIn Ver. 2.0

<210> 1

<211> 1038

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> (1)..(1038)

<400> 1

atg tac aac ggg tgc tgc tgc cgc atc gag ggg gac acc atc tcc cag	48
Met Tyr Asn Gly Ser Cys Cys Arg Ile Glu Gly Asp Thr Ile Ser Gln	
1 5 10 15	
gtg atg ccg ccg ctg ctc att gtg gcc ttt gtg ctg ggc gca cta ggc	96
Val Met Pro Pro Leu Leu Ile Val Ala Phe Val Leu Gly Ala Leu Gly	
20 25 30	
aat ggg gtc gcc ctg tgt ggt ttc tgc ttc cac atg aag acc tgg aag	144
Asn Gly Val Ala Leu Cys Gly Phe Cys Phe His Met Lys Thr Trp Lys	
35 40 45	
ccc agc act gtt tac ctt ttc aat ttg gcc gtg gct gat ttc ctc ctt	192
Pro Ser Thr Val Tyr Leu Phe Asn Leu Ala Val Ala Asp Phe Leu Leu	
50 55 60	
atg atc tgc ctg cct ttt cgg aca gac tat tac ctc aga cgt aga cac	240
Met Ile Cys Leu Pro Phe Arg Thr Asp Tyr Tyr Leu Arg Arg Arg His	
65 70 75 80	
tgg gct ttt ggg gac att ccc tgc cga gtg ggg ctc ttc acg ttg gcc	288
Trp Ala Phe Gly Asp Ile Pro Cys Arg Val Gly Leu Phe Thr Leu Ala	
85 90 95	
atg aac agg gcc ggg agc atc gtg ttc ctt acg gtg gtg gct gcg gac	336
Met Asn Arg Ala Gly Ser Ile Val Phe Leu Thr Val Val Ala Ala Asp	
100 105 110	
agg tat ttc aaa gtg gtc cac ccc cac cac gcg gtg aac act atc tcc	384
Arg Tyr Phe Lys Val Val His Pro His His Ala Val Asn Thr Ile Ser	
115 120 125	

acc cgg gtg gcg gct ggc atc gtc tgc acc ctg tgg gcc ctg gtc atc 432  
 Thr Arg Val Ala Ala Gly Ile Val Cys Thr Leu Trp Ala Leu Val Ile  
 130 135 140

ctg gga aca gtg tat ctt ttg ctg gag aac cat ctc tgc gtg caa gag 480  
 Leu Gly Thr Val Tyr Leu Leu Leu Glu Asn His Leu Cys Val Gln Glu  
 145 150 155 160

acg gcc gtc tcc tgt gag agc ttc atc atg gag tcg gcc aat ggc tgg 528  
 Thr Ala Val Ser Cys Glu Ser Phe Ile Met Glu Ser Ala Asn Gly Trp  
 165 170 175

cat gac atc atg ttc cag ctg gag ttc ttt atg ccc ctc ggc atc atc 576  
 His Asp Ile Met Phe Gln Leu Glu Phe Phe Met Pro Leu Gly Ile Ile  
 180 185 190

tta ttt tgc tcc ttc aag att gtt tgg agc ctg agg cgg agg cag cag 624  
 Leu Phe Cys Ser Phe Lys Ile Val Trp Ser Leu Arg Arg Arg Gln Gln  
 195 200 205

ctg gcc aga cag gct cgg atg aag aag gcg acc cgg ttc atc atg gtg 672  
 Leu Ala Arg Gln Ala Arg Met Lys Lys Ala Thr Arg Phe Ile Met Val  
 210 215 220

gtg gca att gtg ttc atc aca tgc tac ctg ccc agc gtg tct gct aga 720  
 Val Ala Ile Val Phe Ile Thr Cys Tyr Leu Pro Ser Val Ser Ala Arg  
 225 230 235 240

ctc tat ttc ctc tgg acg gtg ccc tcg agt gcc tgc gat ccc tct gtc 768  
 Leu Tyr Phe Leu Trp Thr Val Pro Ser Ser Ala Cys Asp Pro Ser Val  
 245 250 255

cat ggg gcc ctg cac ata acc ctc agc ttc acc tac atg aac agc atg 816  
 His Gly Ala Leu His Ile Thr Leu Ser Phe Thr Tyr Met Asn Ser Met  
 260 265 270

ctg gat ccc ctg gtg tat tat ttt tca agc ccc tcc ttt ccc aaa ttc 864  
 Leu Asp Pro Leu Val Tyr Tyr Phe Ser Ser Pro Ser Phe Pro Lys Phe  
 275 280 285

tac aac aag ctc aaa atc tgc agt ctg aaa ccc aag cag cca gga cac 912  
 Tyr Asn Lys Leu Lys Ile Cys Ser Leu Lys Pro Lys Gln Pro Gly His  
 290 295 300

tca aaa aca caa agg ccg gaa gag atg cca att tcg aac ctc ggt cgc 960  
 Ser Lys Thr Gln Arg Pro Glu Glu Met Pro Ile Ser Asn Leu Gly Arg  
 305 310 315 320

agg agt tgc atc agt gtg gca aat agt ttc caa agc cag tct gat ggg 1008  
 Arg Ser Cys Ile Ser Val Ala Asn Ser Phe Gln Ser Gln Ser Asp Gly  
 325 330 335

caa tgg gat ccc cac att gtt gag tgg cac 1038  
 Gln Trp Asp Pro His Ile Val Glu Trp His  
 340 345

&lt;210&gt; 2

&lt;211&gt; 346

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 2

```

Met Tyr Asn Gly Ser Cys Cys Arg Ile Glu Gly Asp Thr Ile Ser Gln
 1           5           10           15

Val Met Pro Pro Leu Leu Ile Val Ala Phe Val Leu Gly Ala Leu Gly
      20           25           30

Asn Gly Val Ala Leu Cys Gly Phe Cys Phe His Met Lys Thr Trp Lys
 35           40           45

Pro Ser Thr Val Tyr Leu Phe Asn Leu Ala Val Ala Asp Phe Leu Leu
 50           55           60

Met Ile Cys Leu Pro Phe Arg Thr Asp Tyr Tyr Leu Arg Arg Arg His
 65           70           75           80

Trp Ala Phe Gly Asp Ile Pro Cys Arg Val Gly Leu Phe Thr Leu Ala
      85           90           95

Met Asn Arg Ala Gly Ser Ile Val Phe Leu Thr Val Val Ala Ala Asp
 100           105           110

Arg Tyr Phe Lys Val Val His Pro His His Ala Val Asn Thr Ile Ser
 115           120           125

Thr Arg Val Ala Ala Gly Ile Val Cys Thr Leu Trp Ala Leu Val Ile
 130           135           140

Leu Gly Thr Val Tyr Leu Leu Leu Glu Asn His Leu Cys Val Gln Glu
 145           150           155           160

Thr Ala Val Ser Cys Glu Ser Phe Ile Met Glu Ser Ala Asn Gly Trp
      165           170           175

His Asp Ile Met Phe Gln Leu Glu Phe Phe Met Pro Leu Gly Ile Ile
 180           185           190

Leu Phe Cys Ser Phe Lys Ile Val Trp Ser Leu Arg Arg Arg Gln Gln
 195           200           205

Leu Ala Arg Gln Ala Arg Met Lys Lys Ala Thr Arg Phe Ile Met Val
 210           215           220

Val Ala Ile Val Phe Ile Thr Cys Tyr Leu Pro Ser Val Ser Ala Arg
 225           230           235           240

Leu Tyr Phe Leu Trp Thr Val Pro Ser Ser Ala Cys Asp Pro Ser Val
      245           250           255

His Gly Ala Leu His Ile Thr Leu Ser Phe Thr Tyr Met Asn Ser Met
 260           265           270

Leu Asp Pro Leu Val Tyr Tyr Phe Ser Ser Pro Ser Phe Pro Lys Phe
 275           280           285

Tyr Asn Lys Leu Lys Ile Cys Ser Leu Lys Pro Lys Gln Pro Gly His
 290           295           300

Ser Lys Thr Gln Arg Pro Glu Glu Met Pro Ile Ser Asn Leu Gly Arg

```

```

305                      310                      315                      320
Arg Ser Cys Ile Ser Val Ala Asn Ser Phe Gln Ser Gln Ser Asp Gly
                      325                      330                      335

Gln Trp Asp Pro His Ile Val Glu Trp His
                      340                      345

<210> 3
<211> 3251
<212> DNA
<213> Mus musculus

<220>
<221> CDS
<222> (350) .. (1402)

<400> 3
gaaaaagaca aaaccagaaa aagaaaagtc atctccaggg ctcgatctag caacgagtct 60
gtagcatgta tagcgtcgga ccccgagct gcaaccaga aatgtacact cgtgggaaac 120
cgcttgacac ccagagcctg acccagctgc aggcttcaac tctgtagggg acgtgcagct 180
cgtgatccaa gcctaggaga aaggacttgc tgccggcttt catttcctgg ctgaagtttc 240
tctcgtgggt gcagcgctg catcccaggg tgatgaggtt aggggccag ctgctagagg 300
agccctagtg ttcggatagg cagctgtgcc tctgtgccg ccaccttgg atg cca gtc 358
                                   Met Pro Val
                                   1

ctc tct cca act gct atg gac aac ggg tcg tgc tgt ctc atc gag ggg 406
Leu Ser Pro Thr Ala Met Asp Asn Gly Ser Cys Cys Leu Ile Glu Gly
   5                      10                      15

gag ccc atc tcc cag gtg atg cct cct cta ctc atc ctg gtc ttc gtg 454
Glu Pro Ile Ser Gln Val Met Pro Pro Leu Leu Ile Leu Val Phe Val
  20                      25                      30                      35

ctt ggc gcc ctg ggc aac ggc ata gcc ctg tgc ggc ttc tgc ttt cac 502
Leu Gly Ala Leu Gly Asn Gly Ile Ala Leu Cys Gly Phe Cys Phe His
          40                      45                      50

atg aag acc tgg aag tca agc act att tac ctt ttc aac ttg gct gtg 550
Met Lys Thr Trp Lys Ser Ser Thr Ile Tyr Leu Phe Asn Leu Ala Val
          55                      60                      65

gcc gat ttt ctc ctc atg atc tgc tta ccc ctt cgg aca gac tac tac 598
Ala Asp Phe Leu Leu Met Ile Cys Leu Pro Leu Arg Thr Asp Tyr Tyr
          70                      75                      80

ctc aga cgc aga cac tgg att ttt gga gat atc gcc tgt cgc ctg gtc 646
Leu Arg Arg Arg His Trp Ile Phe Gly Asp Ile Ala Cys Arg Leu Val
          85                      90                      95

ctc ttc aag ctg gcc atg aat agg gcc ggg agc att gtc ttc ctc act 694
Leu Phe Lys Leu Ala Met Asn Arg Ala Gly Ser Ile Val Phe Leu Thr
100                      105                      110                      115

```

gtg gtg gct gtg gat agg tat ttc aaa gtg gtc cac ccc cac cat atg 742  
 Val Val Ala Val Asp Arg Tyr Phe Lys Val Val His Pro His His Met  
 120 125 130

gtg aat gcc atc tcc aac cgg act gcc gcc gcc acc gcc tgt gtc ctc 790  
 Val Asn Ala Ile Ser Asn Arg Thr Ala Ala Ala Thr Ala Cys Val Leu  
 135 140 145

tgg act ttg gtc atc ttg ggg act gtg tat ctt ctg atg gag agt cac 838  
 Trp Thr Leu Val Ile Leu Gly Thr Val Tyr Leu Leu Met Glu Ser His  
 150 155 160

ctg tgt gtg cag ggg aca ctg tgg tcc tgt gag agc ttc atc atg gag 886  
 Leu Cys Val Gln Gly Thr Leu Ser Ser Cys Glu Ser Phe Ile Met Glu  
 165 170 175

tca gcc aac ggg tgg cac gat gtc atg ttc cag ctg gag ttc ttc ctg 934  
 Ser Ala Asn Gly Trp His Asp Val Met Phe Gln Leu Glu Phe Phe Leu  
 180 185 190 195

ccc ctg aca atc atc ttg ttc tgc tgg gtc aac gtt gtt tgg agc ctg 982  
 Pro Leu Thr Ile Ile Leu Phe Cys Ser Val Asn Val Val Trp Ser Leu  
 200 205 210

aga cgg agg cag cag ctg acc aga cag gct cgg atg agg agg gcc acc 1030  
 Arg Arg Arg Gln Gln Leu Thr Arg Gln Ala Arg Met Arg Arg Ala Thr  
 215 220 225

cgg ttc atc atg gtg gtg gct tct gtg ttc atc acg tgt tac ctg ccc 1078  
 Arg Phe Ile Met Val Val Ala Ser Val Phe Ile Thr Cys Tyr Leu Pro  
 230 235 240

agc gtg ctg gct agg ctc tac ttc ctc tgg acg gtg ccc act agt gcc 1126  
 Ser Val Leu Ala Arg Leu Tyr Phe Leu Trp Thr Val Pro Thr Ser Ala  
 245 250 255

tgt gac ccc tct gtc cac aca gcc ctc cac gtc acc ctg agc ttc acc 1174  
 Cys Asp Pro Ser Val His Thr Ala Leu His Val Thr Leu Ser Phe Thr  
 260 265 270 275

tac ctg aac agt atg ctg gat ccc ctt gta tat tac ttc tca agc ccc 1222  
 Tyr Leu Asn Ser Met Leu Asp Pro Leu Val Tyr Tyr Phe Ser Ser Pro  
 280 285 290

tgg ctc ccc aaa ttc tac gcc aag ctc aca atc tgc agc ctg aag ccc 1270  
 Ser Leu Pro Lys Phe Tyr Ala Lys Leu Thr Ile Cys Ser Leu Lys Pro  
 295 300 305

aaa cgc cca gga cgc acg aag acg cgg agg tca gaa gag atg cca att 1318  
 Lys Arg Pro Gly Arg Thr Lys Thr Arg Arg Ser Glu Glu Met Pro Ile  
 310 315 320

tgg aac ctc tgc agt aag agc tcc atc gat ggg gca aat cgt tcc cag 1366  
 Ser Asn Leu Cys Ser Lys Ser Ser Ile Asp Gly Ala Asn Arg Ser Gln  
 325 330 335

agg cca tct gac ggg cag tgg gat ctc caa gtg tgt tgaatgccat 1412  
 Arg Pro Ser Asp Gly Gln Trp Asp Leu Gln Val Cys  
 340 345 350

taagacaaac agcccaacaa cgaggcagag aaatgggcaa tgtgagttaa atctgaaggg 1472

tggaggactt gaagatgtcc cctcccactc ttagctgtat ctttctcact caggtagaaa 1532  
tgggatccac cctgcttgac cttttccaga aggttccaaa cgggttggtt gtgtttaaat 1592  
actctgatag caatggtgaa ggggcagcgt gtgagtgtga aggaaaccgt ggggtgtcggg 1652  
ttaggaacta cctggagccc gtgtcgcttt gcatggctga gaaaagcggg atgagcctgg 1712  
ctgggtcttg ttctagctcg gagagagttt acgatctcaa taactcgtcg gtatttcctg 1772  
gactgaaaaa aatagaaact gcactgagtc aatacactta tttccagctg agcgagacct 1832  
tttactgcag gacacccgga cctagccgtt tttttaaatc ttccctgggg agcctccaca 1892  
catttcaagg tttgaacatc cagggtggccc aggagggcag caaaaagaat ctattctaac 1952  
cttgctggcg ccacgatatt ttgctgattt taagtggttt catcctttgt ttttcttttg 2012  
tttgcatctc aaagaagatg ctgagggact tgtccacctg atatcagcta tcgtttctcc 2072  
agtgggaact gagagcctgt ttacggcagc aatgggtgggg ggggggtgct tcctggtatt 2132  
tgggatgggt taattcaagc atggttggtc ttactgctt aatgcatgaa tttgagctga 2192  
aatcctccct tctcaagtct ttgtttaatc cacagtatgt tgtcccacc tgtccagcat 2252  
cctcgtctgt tttgtctttg gtgctgggca ttgaacttg gacctattc gagccaggca 2312  
agcactgacc actgaactgc actcccaaac cccttggtgc ctttttagct gtagcgttgt 2372  
tagccaactt ttgggagaaa gcaaagcact agaggtggca gcaacagttt agctcaatgt 2432  
cctttcgtca gtgtctagac ttctggtcag ccacccgggt ctcctattgg gggcctccct 2492  
caagcacata tgttctccca aatactacc agaattctca cagctaggtg attctgtgaa 2552  
agtccagggt gccctgtcc tggagaagga gaaatagaat ccgtgttaac cttagtccca 2612  
ctttcaagcc acaaaagtgg tgacagccat tcactctctc cagttcccag ggtactctcc 2672  
ccagggaagg gacctgaca tttatgtcta aagacataaa ttagatgctc ctcaagggtg 2732  
tccctgtggc ttcctttgcc agaggtgttg aagcctaggt gcgaaaatca gtctgactgc 2792  
agggtcgttg agaaggctca gtaggtaaac aggtctgccg ccaagccacc aacctgagta 2852  
tccctgagac ccacgtggtg ggagaacacc aaatcccaa ggtggttctg tcctccacct 2912  
atgttcacgc atgcgacat gcacgcatgc gcacgcgcgc acaccaccc accaaataca 2972  
ttaatgcaat aaaaatttta ttggctacac ggtcaagttt gaatcttagt ttaaagtctt 3032  
attagacatg tgctcgtagg gaagacttta ttaacccca ctcagttttg atgttcagca 3092  
gggttaatac tgatgccaaa ggggttgga gcaaattcaa tgactgtaca ctcataagca 3152  
tgagaaatct gttgttccca ggtctgcccga gaagaagacc atgtgcgtgt agtagttgat 3212  
aaataaatag ttgctgaaca actataatcg ctccaaaaa 3251

&lt;210&gt; 4

&lt;211&gt; 351

&lt;212&gt; PRT

&lt;213&gt; Mus musculus

&lt;400&gt; 4

```

Met Pro Val Leu Ser Pro Thr Ala Met Asp Asn Gly Ser Cys Cys Leu
 1           5           10           15

Ile Glu Gly Glu Pro Ile Ser Gln Val Met Pro Pro Leu Leu Ile Leu
      20           25           30

Val Phe Val Leu Gly Ala Leu Gly Asn Gly Ile Ala Leu Cys Gly Phe
      35           40           45

Cys Phe His Met Lys Thr Trp Lys Ser Ser Thr Ile Tyr Leu Phe Asn
      50           55           60

Leu Ala Val Ala Asp Phe Leu Leu Met Ile Cys Leu Pro Leu Arg Thr
      65           70           75           80

Asp Tyr Tyr Leu Arg Arg Arg His Trp Ile Phe Gly Asp Ile Ala Cys
      85           90           95

Arg Leu Val Leu Phe Lys Leu Ala Met Asn Arg Ala Gly Ser Ile Val
      100           105           110

Phe Leu Thr Val Val Ala Val Asp Arg Tyr Phe Lys Val Val His Pro
      115           120           125

His His Met Val Asn Ala Ile Ser Asn Arg Thr Ala Ala Ala Thr Ala
      130           135           140

Cys Val Leu Trp Thr Leu Val Ile Leu Gly Thr Val Tyr Leu Leu Met
      145           150           155           160

Glu Ser His Leu Cys Val Gln Gly Thr Leu Ser Ser Cys Glu Ser Phe
      165           170           175

Ile Met Glu Ser Ala Asn Gly Trp His Asp Val Met Phe Gln Leu Glu
      180           185           190

Phe Phe Leu Pro Leu Thr Ile Ile Leu Phe Cys Ser Val Asn Val Val
      195           200           205

Trp Ser Leu Arg Arg Arg Gln Gln Leu Thr Arg Gln Ala Arg Met Arg
      210           215           220

Arg Ala Thr Arg Phe Ile Met Val Val Ala Ser Val Phe Ile Thr Cys
      225           230           235           240

Tyr Leu Pro Ser Val Leu Ala Arg Leu Tyr Phe Leu Trp Thr Val Pro
      245           250           255

Thr Ser Ala Cys Asp Pro Ser Val His Thr Ala Leu His Val Thr Leu
      260           265           270

Ser Phe Thr Tyr Leu Asn Ser Met Leu Asp Pro Leu Val Tyr Tyr Phe
      275           280           285

```

Ser Ser Pro Ser Leu Pro Lys Phe Tyr Ala Lys Leu Thr Ile Cys Ser  
 290 295 300

Leu Lys Pro Lys Arg Pro Gly Arg Thr Lys Thr Arg Arg Ser Glu Glu  
 305 310 315 320

Met Pro Ile Ser Asn Leu Cys Ser Lys Ser Ser Ile Asp Gly Ala Asn  
 325 330 335

Arg Ser Gln Arg Pro Ser Asp Gly Gln Trp Asp Leu Gln Val Cys  
 340 345 350

<210> 5  
 <211> 1668  
 <212> DNA  
 <213> Rattus norvegicus

<220>  
 <221> CDS  
 <222> (365) .. (1417)

<400> 5  
 tggtagcgct gcaggtaccg gtccggaatt cccgggtcga cccacgcgtc cgcaaaacta 60  
 gaaaagcaaaa atcgtcccca ggggtggacc cagcgacaag tctgctgcgt ggctggcatc 120  
 agacccccaa gctgcagcct ggcaatgtac gcttttggaa aactgctctc gcctcagagc 180  
 ctgaccacagc tgcaggcttc acctctgtag gggacatgca gcttgtgatc caggctgagg 240  
 agaaaggacc tgctgtcggc ttctatttcc tgactgaagt tgggtgatga ggtaggggc 300  
 ccagctgcca aggggaacca tagtgttcag ataggcagct gtgcctttgt gtcagccacc 360  
 tagg atg ctc ttc ctc tct ccg agt gct atg gac aac ggg tcg tgc tgt 409  
 Met Leu Phe Leu Ser Pro Ser Ala Met Asp Asn Gly Ser Cys Cys  
 1 5 10 15

ctc atc gag ggg gaa ccc atc acc cag gta atg cca cct tta ctc atc 457  
 Leu Ile Glu Gly Glu Pro Ile Thr Gln Val Met Pro Pro Leu Leu Ile  
 20 25 30

ctg gcc ttc ctg ctt gga gcc ctg ggc aac ggc cta gcc ctg tgt ggt 505  
 Leu Ala Phe Leu Leu Gly Ala Leu Gly Asn Gly Leu Ala Leu Cys Gly  
 35 40 45

ttc tgc ttt cac atg aag acc tgg aag tcg agc act att tac ctt ttc 553  
 Phe Cys Phe His Met Lys Thr Trp Lys Ser Ser Thr Ile Tyr Leu Phe  
 50 55 60

aac ttg gct gta gcc gat ttt ctc ctc atg atc tgc cta ccc ctt cgg 601  
 Asn Leu Ala Val Ala Asp Phe Leu Leu Met Ile Cys Leu Pro Leu Arg  
 65 70 75

aca gac tac tac ctc aga cgt agg cat tgg att ttg ggg gat att ccc 649  
 Thr Asp Tyr Tyr Leu Arg Arg Arg His Trp Ile Leu Gly Asp Ile Pro  
 80 85 90 95

tgc cgc ctg gtc ctc ttc atg ctg gcc atg aat agg gcc gga agc att 697  
 Cys Arg Leu Val Leu Phe Met Leu Ala Met Asn Arg Ala Gly Ser Ile



100										105										110										
gtc ttc ctc act gtg gtg gcc gtg gac agg tat ttc aaa gtg gtc cac	745																													
Val Phe Leu Thr Val Val Ala Val Asp Arg Tyr Phe Lys Val Val His																														
115 120 125																														
ccc cac cat atg gtg aac gcc atc tcc aat cgg act gca gct gcc atc	793																													
Pro His His Met Val Asn Ala Ile Ser Asn Arg Thr Ala Ala Ala Ile																														
130 135 140																														
gtc tgt gtc ctc tgg act ttg gtc atc ttg ggg act gtg tat ctt ctg	841																													
Val Cys Val Leu Trp Thr Leu Val Ile Leu Gly Thr Val Tyr Leu Leu																														
145 150 155																														
atg gag agt cac ctg tgt gtg cgg ggg atg gtg tca tct tgt gag agc	889																													
Met Glu Ser His Leu Cys Val Arg Gly Met Val Ser Ser Cys Glu Ser																														
160 165 170 175																														
ttc atc atg gag tca gcc aac ggg tgg cac gat atc atg ttc cag ctg	937																													
Phe Ile Met Glu Ser Ala Asn Gly Trp His Asp Ile Met Phe Gln Leu																														
180 185 190																														
gag ttc ttc ctg ccc ctg acc atc atc ttg ttc tgc tcc ttc aaa gtt	985																													
Glu Phe Phe Leu Pro Leu Thr Ile Ile Leu Phe Cys Ser Phe Lys Val																														
195 200 205																														
gtt tgg agc ctg aga cag agg caa cag ctg acc aga cag gct cgg atg	1033																													
Val Trp Ser Leu Arg Gln Arg Gln Gln Leu Thr Arg Gln Ala Arg Met																														
210 215 220																														
agg agg gcc acc cgg ttc atc atg gtg gtg gct tcc gtg ttc atc acg	1081																													
Arg Arg Ala Thr Arg Phe Ile Met Val Val Ala Ser Val Phe Ile Thr																														
225 230 235																														
tgt tac ctg ccc agc gtg ttg gcg agg ctc tac ttc ctc tgg acg gtg	1129																													
Cys Tyr Leu Pro Ser Val Leu Ala Arg Leu Tyr Phe Leu Trp Thr Val																														
240 245 250 255																														
ccc tcc agt gct tgt gac ccc tct gtc cac ata gct ctc cat gtc acc	1177																													
Pro Ser Ser Ala Cys Asp Pro Ser Val His Ile Ala Leu His Val Thr																														
260 265 270																														
ctg agt ctc acc tac ctg aac agc atg ctg gac cct ctt gtg tac tac	1225																													
Leu Ser Leu Thr Tyr Leu Asn Ser Met Leu Asp Pro Leu Val Tyr Tyr																														
275 280 285																														
ttt tca agc ccc tcg ttc ccc aaa ttc tac gcc aag ctc aaa atc cgc	1273																													
Phe Ser Ser Pro Ser Phe Pro Lys Phe Tyr Ala Lys Leu Lys Ile Arg																														
290 295 300																														
agc ttg aaa ccc aga cgc cca gga cgc tcg cag gca cgg agg tcg gaa	1321																													
Ser Leu Lys Pro Arg Arg Pro Gly Arg Ser Gln Ala Arg Arg Ser Glu																														
305 310 315																														
gag atg cca att tcg aat ctc tgt cgt aag agt tcc acc gat gtg gta	1369																													
Glu Met Pro Ile Ser Asn Leu Cys Arg Lys Ser Ser Thr Asp Val Val																														
320 325 330 335																														
aat agt tcc cag agg ccg tct gac ggg cag tgg ggt ctc caa gtg tgt	1417																													
Asn Ser Ser Gln Arg Pro Ser Asp Gly Gln Trp Gly Leu Gln Val Cys																														
340 345 350																														

tgaatgccat gaagacaaat ggcccagcag caaagcagag acctgggcaa ctgtgagtta 1477  
aatctgaagg gtgagggaact tgaaaaatga cagccccccc ccccgcccca cccgcccgcc 1537  
cgccccgctc tttctcagct gtgtctttct cactcaagta gaagcaaaat ctaaaaaaaaa 1597  
aaaaaaaaaa aaaaaaaaaa agggcgcccg ctctagagga tccaagctta cgtacgcgtg 1657  
catgcgacgt c 1668

<210> 6

<211> 351

<212> PRT

<213> Rattus norvegicus

<400> 6

Met	Leu	Phe	Leu	Ser	Pro	Ser	Ala	Met	Asp	Asn	Gly	Ser	Cys	Cys	Leu	1	5	10	15
Ile	Glu	Gly	Glu	Pro	Ile	Thr	Gln	Val	Met	Pro	Pro	Leu	Leu	Ile	Leu	20	25	30	
Ala	Phe	Leu	Leu	Gly	Ala	Leu	Gly	Asn	Gly	Leu	Ala	Leu	Cys	Gly	Phe	35	40	45	
Cys	Phe	His	Met	Lys	Thr	Trp	Lys	Ser	Ser	Thr	Ile	Tyr	Leu	Phe	Asn	50	55	60	
Leu	Ala	Val	Ala	Asp	Phe	Leu	Leu	Met	Ile	Cys	Leu	Pro	Leu	Arg	Thr	65	70	75	80
Asp	Tyr	Tyr	Leu	Arg	Arg	Arg	His	Trp	Ile	Leu	Gly	Asp	Ile	Pro	Cys	85	90	95	
Arg	Leu	Val	Leu	Phe	Met	Leu	Ala	Met	Asn	Arg	Ala	Gly	Ser	Ile	Val	100	105	110	
Phe	Leu	Thr	Val	Val	Ala	Val	Asp	Arg	Tyr	Phe	Lys	Val	Val	His	Pro	115	120	125	
His	His	Met	Val	Asn	Ala	Ile	Ser	Asn	Arg	Thr	Ala	Ala	Ala	Ile	Val	130	135	140	
Cys	Val	Leu	Trp	Thr	Leu	Val	Ile	Leu	Gly	Thr	Val	Tyr	Leu	Leu	Met	145	150	155	160
Glu	Ser	His	Leu	Cys	Val	Arg	Gly	Met	Val	Ser	Ser	Cys	Glu	Ser	Phe	165	170	175	
Ile	Met	Glu	Ser	Ala	Asn	Gly	Trp	His	Asp	Ile	Met	Phe	Gln	Leu	Glu	180	185	190	
Phe	Phe	Leu	Pro	Leu	Thr	Ile	Ile	Leu	Phe	Cys	Ser	Phe	Lys	Val	Val	195	200	205	
Trp	Ser	Leu	Arg	Gln	Arg	Gln	Gln	Leu	Thr	Arg	Gln	Ala	Arg	Met	Arg	210	215	220	
Arg	Ala	Thr	Arg	Phe	Ile	Met	Val	Val	Ala	Ser	Val	Phe	Ile	Thr	Cys				

```

225                230                235                240
Tyr Leu Pro Ser Val Leu Ala Arg Leu Tyr Phe Leu Trp Thr Val Pro
                245                250                255
Ser Ser Ala Cys Asp Pro Ser Val His Ile Ala Leu His Val Thr Leu
        260                265                270
Ser Leu Thr Tyr Leu Asn Ser Met Leu Asp Pro Leu Val Tyr Tyr Phe
        275                280                285
Ser Ser Pro Ser Phe Pro Lys Phe Tyr Ala Lys Leu Lys Ile Arg Ser
        290                295                300
Leu Lys Pro Arg Arg Pro Gly Arg Ser Gln Ala Arg Arg Ser Glu Glu
305                310                315                320
Met Pro Ile Ser Asn Leu Cys Arg Lys Ser Ser Thr Asp Val Val Asn
        325                330                335
Ser Ser Gln Arg Pro Ser Asp Gly Gln Trp Gly Leu Gln Val Cys
        340                345                350

```

```

<210> 7
<211> 15
<212> PRT
<213> Human immunodeficiency virus type 1

```

```

<400> 7
Gly Gly Gly Gly Tyr Gly Arg Lys Lys Arg Arg Gln Arg Arg Arg
 1                5                10                15

```

```

<210> 8
<211> 11
<212> PRT
<213> Artificial Sequence

```

```

<220>
<223> Description of Artificial Sequence: internalizing
        domain derived from HIV tat protein

```

```

<400> 8
Tyr Gly Arg Lys Lys Arg Arg Gln Arg Arg Arg
 1                5                10

```

```

<210> 9
<211> 25
<212> DNA
<213> Artificial Sequence

```

```

<220>
<223> Description of Artificial Sequence: PCR primer

```

```

<400> 9
aagaggacca ggccggcaggg aatat

```

25

```

<210> 10

```

<211> 25  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: PCR primer

<400> 10  
tatcccccaa aatccaatgc ctacg 25

<210> 11  
<211> 24  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: forward primer

<400> 11  
cgggcaggtg ggtgatgagg ttag 24

<210> 12  
<211> 23  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: reverse primer

<400> 12  
gctgctgggc catttgtctt cat 23

<210> 13  
<211> 21  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence:  
oligonucleotide probe

<400> 13  
tgctgtctca tcgaggggga a 21

<210> 14  
<211> 21  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence:  
oligonucleotide probe

<400> 14  
gaatagggcc ggaagcattg t 21

<210> 15  
<211> 23  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: PCR primer

<400> 15  
cctcctcatc cgagcctgtc tgg 23

<210> 16  
<211> 25  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: PCR primer

<400> 16  
cctttgtgtc agccacctag gatgc 25

<210> 17  
<211> 19  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: PCR primer

<400> 17  
ttcacgttgg ccatgaaca 19

<210> 18  
<211> 18  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: PCR primer

<400> 18  
aaataacctgt ccgcagcc 18

<210> 19  
<211> 24  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence:  
oligonucleotide probe

<400> 19  
ccgtaaggaa cacgatgctc ccgg 24

<210> 20  
<211> 15  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: PCR primer

<400> 20  
gtcgacggcg agccc

15

<210> 21  
<211> 19  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence: PCR primer

<400> 21  
tctttgggac cttgtctgc

19

<210> 22  
<211> 22  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Description of Artificial Sequence:  
oligonucleotide probe

<400> 22  
tgggccgcgt ctcctttgag ct

22